Evaluation of Sensitivity and Uncertainty Analysis Methods in a Quality Assessment Framework with Application to Environmental and Business Statistics

Lot 14 of EUROSTAT SUP COM 1997
JRC reference: Administrative Arrangement 13592-1998-01 A1CA ISP LU

FINAL REPORT

June 1999

Prepared by JRC-ISIS-SAIE-UASA, June, 17-23, 1999
European Commission
Joint Research Centre
Institute for Systems, Informatics and Safety
T.P. 361 – 21020 Ispra (VA) – ITALY
e-mail: andrea.saltelli@jrc.it

1 EXECUTIVE SUMMARY

The main objectives of this research project demanded by EUROSTAT are:

- to ascertain the improvement that can be achieved using Sensitivity Analysis (SA) and Uncertainty Analysis (UA) in measuring the quality of statistics,
- to test recent UA, SA methodologies for the management of uncertainties on case studies relevant to EUROSTAT' needs, and
- to contrast those results with those obtainable with standard statistical methods.

An heuristic approach is taken: two specific test cases are worked out, one in the area of environmental statistics, and another in the area of business statistics. All inference about the use of UA and SA is made based on the test cases, rather than on general considerations.

The **first test case** focused on global sensitivity analysis as a quality assurance tool in environmental policy modelling.

It illustrates an application of Sensitivity Analysis (SA) in the context of quality assessment in environmental statistics, showing that SA is an indispensable tool for the management of uncertainties in models used for environmental policy and for the transparency of the models themselves.
The case study originates from EUROSTAT. We illustrate the role that global sensitivity analysis can play as a quality assurance tool for the optimisation of the uncertainties on a worked example related to the use of environmental indicators for policy-making. Uncertainty Analysis (UA) is employed to test the robustness of the underlying model, where the uncertain factors (i.e., input variables and model parameters) are explored over their range of variation. In this specific study, UA helps us investigate whether or not the underlying indicator-based model allows a given policy decision to be taken.

Then the Extended FAST (a method for sensitivity analysis) is used to reveal to what extent the uncertain model factors affect the model response. In this study, the factors are clustered according to different logical groups, and the Extended FAST allows us to rank quantitatively the groups of factors according to their influence on the output uncertainty (taking into account all possible interactions occurring between the groups).
The information provided by a global SA can indeed be used in decision making to try to balance the influence of the various uncertain elements. The global SA procedure could be performed iteratively until the optimal situation is reached, which is ideally that corresponding to the case where the elements of the chain are more or less equally influential or otherwise optimal once the cost of minimising the various uncertainties is taken into account.

Figure 3.3 (page 45) shows how the pressure to decision index (a function of several indicators) is uncertain, given the uncertainty in both the selection of a given set of indicators and the underlying input data (mostly from CORINAIR). Further, global sensitivity analysis shows that the former (the set of indicators) has more impact than the latter (the uncertainty in data). Hence for this particular application, artificially based on whether incineration or landfill must be preferred for Austria in 1994, one should not invest in collecting better data but in achieving consensus among the experts on a reference set of indicators.

The **second test case** concerns an application of uncertainty and sensitivity analysis to signal extraction in economic time series.

A common practice in official statistics and applied macroeconomics is to focus the analysis on some unobserved movements like trend and seasonally adjusted series. Several methods are available for trend extraction and seasonal adjustment. With model-based methods, the estimation of the unobserved components is carried out after selection of a model and maximum likelihood estimation of the parameters. When several models describe the data dynamics in a statistically correct way, an information criterion is

typically used to select a single specification. Given the data, the chosen model and the parameter estimates, the unobserved components are computed by signal extraction either via Wiener-Kolmogorov or Kalman filtering.

Such unobserved component estimators can be seen as conditional on a model specification and on the estimated parameters values. In practice, however, time series models are only description of data, so different models may describe the data equally well. A first problem for an official statistician who needs to produce trends and seasonally adjusted series is thus: within the class of the acceptable models, how sensitive are the unobserved components estimators to the model choice? Furthermore, for a given model specification, the collection of new observations may slightly modify the parameter estimates. Consequently, the unobserved components estimators would present some instability due to the model coefficients update. A second problem is thus: for a given model, how sensitive are the unobserved components estimators to changes in model parameters?

The example worked here assesses the uncertainty in the estimators of the components due to the uncertainty of model specification and of model parameters, and applies sensitivity analysis techniques in order to isolate the sources of uncertainty, which are the most influential for characterising the trend, seasonal and short-term movements in time series. Our study can be distinguished from previous published works as it addresses the problem of model and parameter uncertainty in the unobserved component models framework, and as it applies the sensitivity analysis techniques presented in this handbook.

Focusing on trend, seasonal and irregular movements in an economic time series we develop an in-depth analysis of the effects of parameter uncertainty. Next, for a given range of models, we discuss how sensitivity to model specification can be evaluated. The analysis puts together three main tools: the model-based signal extraction techniques; a Bayesian framework for characterising the distribution of model parameters and for assigning probability to time series models; and variance-based sensitivity analysis technique.

One important conclusion of the exercise is that sensitivity analysis can be seen as another way of looking to model parsimony, complementary to information criteria like the Bayesian information criterion.

Further investigation of the role of sensitivity analysis in business statistics.

One of the typical problems tackled in econometrics is forecasting the future value of an endogenous variable given the past values of a number of exogenous ones. In the quite rare case of deterministic exogenous variables, the forecast is said to be unconditional, i.e. it is not conditioned by the PDF of the exogenous variables, given their deterministic character. Unfortunately, in most econometric applications exogenous variables must be forecasted too, so the forecast becomes conditioned by the PDF of the prediction of the exogenous variables. In Bayesian terms, any forecast based on a model that involves stochastic variables and/or parameters (from now on input factors) is conditioned by the probability distribution of the input factors. The stochastic nature of input factors has two effects on the reliability of forecasts. First, the variance of the estimate is higher since the variance of the inputs increases the variance of the conditional expectations. Second, the choice of the most reliable forecasting model is far more difficult than in the unconditional case. A model that shows a good fit in terms of unconditional forecasting may show a poor performance in the conditional case, because of the variance introduced by the input factors. Often it may be difficult to screen among several models with traditional econometric tests, because the PDF of the forecast in most econometric applications is not normally distributed. In particular it might be difficult to identify the relation between the variance of conditional expectation and the variance of any single input factor.

An application is given in the present report to show how variance-based sensitivity analysis could improve econometric model screening and forecast reliability. Sobol' Method is used to analyse the sensitivity of the conditional forecast of two classical case studies:
(1) Solow's classic (1957) study of the technical change in U.S. Economy and
(2) the consumption function of Klein's (1950) Model 1.

The applications showed that

- SA quantifies the share of conditional variance that can be accounted to any input factor Monte Carlo based tests are not constrained by the normality of the PDF of the outputs
- SA methods are of great help in model screening, as they provide an assessment of how "balanced" a model calibration is given the selected model, the available data and the set of parameters to be calibrated.

The report also discusses the similarities and differences between the statistical concept of **robustness** and sensitivity analysis, by applying SA to a classical example from Huber (see section 2.3).

The list given in section 1.1 refers to the articles that have been (or will be) published in the framework of Lot 14 of EUROSTAT SUP COM 1999, and hence bearing an explicit acknowledgement to the Lot 14 project.

1.1 List of publications

Saltelli, A., Tarantola, S. Campolongo, F., Sensitivity analysis as an ingredient of modelling, forthcoming, *Statistical Science*, ~ November 2000.

Saltelli, A., 1999, Sensitivity analysis. Could better methods be used?, *Journal of Geophysical Research*, **104**(D3), 3789-3793.

Hamby D.M. and Tarantola S., 1999, Exploring sensitivity analysis techniques for the assessment of an environmental transport model, Proceedings of ESREL 99, Garching, Munich (D), Schuëller and Kafka (Eds.), Balkema, Rotterdam,1193-119899.

Saltelli A. ,1999, The role of sensitivity analysis. Transparency, relevance, robustness and parsimony in modelling, to appear on the Proceedings of ESREL 99, Garching Munich (D), September 1999.

A.Saltelli , C.Planas, S.Tarantola, F.Campolongo, 1999, The concept of sensitivity with applications to official statistics, *Research in Official Statistics*, **2**, 47-77.

Campolongo, F., Tarantola, S., and Saltelli, A., 1999, Tackling quantitatively large dimensionality problems, *Computer Physics Communications*, **117**, 75-85.

S. Tarantola, M. Puolamaa, J. Jesinghaus, A. Saltelli, 1998, Global Sensitivity Analysis: A Tool For Environmental Modelling Quality Assessment, New Techniques and Technologies for Statistics, proceedings of New Techniques and Technologies for Statistics , 509-514, November 1998, Sorrento - Italy

S. Tarantola, J. Jesinghaus, M. Puolamaa, 1999, Global sensitivity analysis: a quality assurance tool in environmental policy modelling, Mathematical and Statistical Methods for Sensitivity Analysis, John Wiley and Sons, Probability and Statistics series, forthcoming.

C. Planas, R. Depoutot, 1999, Sensitivity analysis as a tool for time series modelling, Mathematical and Statistical Methods for Sensitivity Analysis, John Wiley and Sons, Probability and Statistics series, forthcoming.

2 AN INTRODUCTION TO SENSITIVITY ANALYSIS

2.1 Sensitivity analysis: a brief overview

The origin of Sensitivity Analysis can be tracked from the theory of Design of Experiments (DOE), which was first introduced by Fisher (1935) in the context of physical experimentation. Assume an experiment has to be performed which involves two quantitative variables, temperature (T) and concentration (C), and a single qualitative variable, catalyst (K). The response is a chemical yield, y(T,C,K). The goal of DOE is to measure the "effects" of this three variables, T, C, and K, on the response.

To this purpose the analyst needs to design "a priori" an experiment able to highlight the relationship between the response and the inputs. The most natural way to design an experiment is to use the one-factor-at-a-time approach (OAT, Daniel, 1958, 1973) which varies the variable of interest one at a time, with the remaining factors held constant to their nominal values. However, the result of an OAT depends on the nominal values used for the other variables, and the behaviour of the response function is described only locally in the input space. That local focus can be related to the physical nature of the experiment. Due to the present day computing facilities, simulation experiments have become a powerful supplement to physical studies. In particular, full representation of uncertainties is possible. The consideration of simulation models and the explicit description of the uncertainty around all input variables characterise global sensitivity analysis.

Sensitivity analysis can thus be seen as the modern evolution of DOE. In short,

Definition: Sensitivity Analysis is the study of how the variation in the output of a model can be apportioned to different sources of variation.
Hence sensitivity analysis is closely related to the concepts of uncertainty and of model.

Originally (Tomovic and Vukobratovic, 1972), SA was created to deal simply with uncertainties in the input variables and model parameters. Over the course of time the ideas have been extended to incorporate model conceptual uncertainty, i.e. uncertainty in model structures, assumptions and specifications (Helton and Burmaster, 1996, Draper et al., 1999). For statisticians, a natural description of uncertainties in variables, parameters and around models is given by the Bayesian framework (see for example Draper, 1995, Kass and Raftery, 1995).

Models are developed to approximate or mimic systems and processes of different natures (e.g. physical, environmental, social, or economic), and of varying complexity. Simulation modelling is often used to mimic difficult or impossible experiments.
This mimic is splendidly highlighted by Rosen's formalisation of the modelling process, (Figure 2.1, Rosen, 1991).

According to Rosen, who refers back to Aristotile's cathegorisation of causality, the world, driven by material and efficient cause, and the model, driven by formal causality, are linked via the process of "encoding" from world to model and "decoding" from model to world. While inside "world" and inside "model" causality reigns, encoding and decoding are not themselves entailed by anything, i.e. they are the objects of the modeller craftsmanship. Yet those two activities are the essence and the purpose of the modelling process: one writes a model in the hope that the decoding operation will provide insight on the world. This is only possible if the uncertainty in the information provided by the model (the substance of use for the decoding exercise) is carefully apportioned to the uncertainty associated with the encoding process.

Rosen's formalisation also helps us to distinguish physical modelling from statistical modelling. In this latter, there are encoded (fixed) rules for the selection of a particular statistical model given the data. The concept of data in statistics is also different from that in physics. In statistics, the sample provides the analyst with a closed universe, while in natural systems the universe must be taken as open, and the encoding is to a large extent the process of deciding what part of it to include in the formalisation.

Another very general definition of sensitivity analysis can be stated as:

Definition: Sensitivity Analysis studies the relationships between information flowing in and out of the model.

SA is hence part of model building. It is used to improve the understanding of the model and to increase the confidence in its predictions. It shows how the model response variables react to changes in the inputs, where by input or factors it is intended data, model structure and model parameters. SA is thus closely linked to uncertainty analysis (UA), which aims to quantify the overall uncertainty associated with the response as a result of uncertainties in the model input.

Modellers may conduct SA to determine

(a) the model resemblance with the process under study,
(b) the quality of model definition,
(c) factors that mostly contribute to the output variability
(d) the region in the space of input factors for which the model variation is maximum
(e) optimal regions within the space of factors for use in a subsequent calibration study
(f) interactions between factors.

Points (a) and (f) are considered in a number of areas. An example can be found in chemistry where SA can evaluate the strength of the relation between kinetic or thermodynamic variables hidden in a complex reaction pathway, and some measurable outputs of the reaction system itself (see Rabitz, 1989). Point (b) is mainly of statistical concern, as will be illustrated in case study 2 (see also Young, 1999). Point (c) has several implications. It gives directions for focusing additional research on the measurement of the relevant factor, as will be underlined in test case 1. Another application is model reduction: SA indicates which model parameters can be discarded because insignificant. The classical t-test used in regression analysis can be seen as a sensitivity analysis tool. The first test case also illustrates how sensitivity analysis becomes an instrument to investigate the relevance of a model for a given task (see also Beck et al., 1997) , and hence to increase the transparency of the model-based decisions (see also Oreskes et al., 1994; Saltelli, Campolongo and Tarantola, 1999). Point (d) is relevant in control theory (see Spear, 1970).
In an environmental study, Pastres et al. (1997) discuss the use of SA in a calibration exercise, as mentioned in point (e).

The list (a to f) is far from being exhaustive. Other applications of SA can be found in reliability engineering, for instance for investigating the changes in the fault tree for a plant which derive from changes in the design components. In economics, the concept of elasticity can be seen as a local sensitivity analysis. See also the reviews by Turanyi (1990), Janssen et al. (1990), Helton (1993), Goldsmith (1998).

In general, UA and SA are performed jointly by executing the model repeatedly for combination of factor values sampled with some probability distribution.

The following steps can be listed (see Figure 2.2):

- specify the target function and select the input of interest
- assign a distribution function to the selected factors
- generate a matrix of inputs with that distribution(s) through an appropriate design
- evaluate the model and compute the distribution of the target function
- select a method for assessing the influence or relative importance of each input factor on the target function.

The first four points are related to UA since they quantify the variation in the model response. SA complements UA by apportioning the uncertainty in the target function to the inputs. A typical representation of the SA results would be a pie chart (Figure 2.2) that partitions the variance of the output. Not all methods rely on the assumptions necessary for that representation.

Different SA methods have different properties, computational cost and application range. **Local sensitivity analysis** concentrates on local impacts of the factors, measured through partial derivatives of the target function. Although examples of successful applications can be found in chemistry, it is not suited for problems involving sizeable uncertainty in the inputs. In dealing with models that are computationally expensive to evaluate and have a large number of input factors, **screening experiments** can be used to identify the factors subset that controls most of the output variability (with low computational effort). This is based on the authors' experience that the influence of factors in models is distributed as income in nations, i.e. follow Pareto's law, with few very influential factors and a majority of non influential ones. As a drawback, these "economical" methods tend to provide qualitative sensitivity measures, i.e. they rank the input factors in order of importance, but do not quantify how much a given factor is more important than another. There is clearly a trade off between computational cost and information.

In contrast with local analysis, **global sensitivity analysis** considers the full range of variation of the input parameters along their joint distribution. Because the factors are varied simultaneously, this involves a multidimensional averaging. Global SA techniques have been discussed by Cukier et al., (1978), Iman and Helton, (1988), Sobol', (1990), Mc Kay 1995, Saltelli, Tarantola and Chan 1999 and Saltelli, Tarantola and Campolongo, 1999, among others.

2.2 Global sensitivity analysis

In this project global, variance-based methods have been used. These are quantitative since they compute the contribution to the total output variance that each input (or group of) is accounting for. For a given factor x_i, the first-order sensitivity measure can be expressed as:

$$S_i = V\left(E\left(Y \mid x_i\right)\right) \Big/ V \qquad (1)$$

where V is the unconditioned variance of the target function Y, the inner expectation in the numerator is taken by fixing factor x_i to some fixed value x_i^*, and the outer variance is over all possible values for x_i^*. Expression (1) is known as importance measure (see Archer et al., 1997). For independent factors, the importance measure is equal to a first order sensitivity index of Sobol' (see Sobol', 1990).

Sobol' approach is based on the analysis of variance (ANOVA), i.e. the output variance is univocally decomposed in orthogonal terms of increasing dimensionality. For example, for a model with three factors the total variance V is decomposed as

$$V = V_1 + V_2 + V_3 + V_{12} + V_{13} + V_{23} + V_{123}\,. \qquad (2)$$

The first-order term quantifies the effect of each single factor; it corresponds to the numerator of (1). The second-order terms explain the effect due to the interaction between a given pair of factors that is not amenable to the linear combination of the effects due to each of them. The third-order is similarly derived. For terms of order higher than one, the equivalence with variances of conditional expectations can be found in Archer et al. (1997). These partial variances are then normalised to the total variance to obtain the sensitivity indices as in (1). Among the advantages of sensitivity indices is their completeness, i.e. their capacity to appreciate all the interaction effects, which are significant especially for non-linear, non-additive models. As a result, the indices are quantitative, because 100% of the output variance is accounted for in the decomposition. This can be contrasted with regression analysis applied to a model output: the regression coefficients can be used as a measure of sensitivity, but the fraction of the model variance taken into consideration is only that accounted for by the regression model, i.e. a fraction equal to the model coefficient of determination R_Y^2; see Saltelli and Homma, (1992).

The apparent drawback of variance based methods is the so-called curse of dimensionality, which is palpable when the number of factors becomes large: the number of terms in the decomposition (2) grows exponentially with the number of factors. This has led to the use of the total sensitivity indices (Sobol', 1990; Homma and Saltelli, 1996). In the same three-factor case, the total sensitivity indices are defined as:

$$
\begin{aligned}
S_{T_1} &= S_1 + S_{12} + S_{13} + S_{123} \\
S_{T_2} &= S_2 + S_{21} + S_{23} + S_{213} \\
S_{T_3} &= S_3 + S_{31} + S_{32} + S_{312}
\end{aligned}
\qquad (3)
$$

For a model with *k* factors there are only *k* total sensitivity indices against the ($2^k - 1$) terms in Equation 2. Each S_{Ti} can be computed independently from the summands that compose it (you do not need S_{13} to evaluate S_{T_1}). Three alternative ways to calculate global sensitivity indices are currently available: the method proposed by Sobol' 1990, based on multidimensional integrals that are computed via Monte Carlo; the extended version of the Fourier Amplitude Sensitivity Test (FAST, Saltelli, Tarantola and Chan, 1999); and a two-step procedure proposed by Rabitz et al. (1999) that involves the so-called cut-high dimensional model representations (cut-HDMR).

In computing the indices, if *R* is the number of replications used to simulate the conditioning argument x_i, and a similar sample of size *R* is used for the complementary set x_{-i} (in Equation (1)), then one needs to evaluate the output a total of R^2 times (test case 2). When the factors are independent (test case 1), a much faster convergence can be achieved by either Sobol' or FAST methods (order of $2 \times R$, see e.g. Saltelli et al., 1993).

A collection of articles on the domain can be found in some recent special issues and proceedings (Journal of Statistical Computation and Simulation (1997), Reliability Engineering and System Safety (1997), Computer Physics Communications (1999), and Journal of Multicriteria Decision Analysis, (1999), Chan et al., (1998)).

Some similarities exist between global SA, the statistical methods of Monte Carlo filtering (Rose et al., 1991, Fedra, et al., 1981), and "Generalised sensitivity analysis" (Young, et al., 1996). In Monte Carlo filtering (that can be thought of as of a generalisation of the calibration concept) one maps the input factors into an output space using plain Monte Carlo. Subsequently, the output sample is censored of the non-plausible or impossible realisations, and the censored set is mapped back into the input factors. The resulting subspace is taken as the result of the calibration, and no attempt is made to identify in this subset an optimum point. In generalised sensitivity analysis one partitions the realisations of the output into acceptable and unacceptable. Again this is then mapped back into the input factors, and for each factor one uses statistics of the Kolmogorov - Smirnov family to tell the two sub-samples for that factor apart. When the test is significant, this implies that the given factor is an influential one. All the above is based on simulation data (although actual observation are used to judge upon the acceptability of the model realisations). Statisticians work as a rule with actual data.

A statistical procedure that resemble the above when using actual data is the bootstrapping of the modelling process (Chatfield, 1993). Here the sample is obtained by re-sampling the given set (as opposed to simulating it), but even in this case one tries to make inference on the stability (robustness) of the estimates: is the same model consistently selected? What is the range of variation in model parameters? Is there any factor more responsible than others of the observed variability? The second test case implements this kind of analysis.

2.3 Sensitivity analysis and the concept of robustness

In the project we have addressed the question of the difference between the concepts of sensitivity and of robustness, which is a classical concept in Statistics. Statisticians are mostly interested in "distributional robustness", intended as invariance in case of small deviations from the assumptions about the underlying distribution assumed for the data (Huber, 1981). The classical example is how to devise estimators that are robust if the data deviate slightly from normality. Often this can be rephrased as "robustness with respect to outliers". Thus robustness is mainly aimed at analysing the efficiency of a given estimator in case of departures from assumptions, in particular about the distribution of the data.

The difference between the two concepts can be highlighted using a pedagogical example taken from Huber (1981, p.2). Some observed data $x_t, t = 1,\ldots,n$ are supposed to be distributed as a mixture of normal distributions:

$$F(x_t) = (1-\varepsilon)N(0,1) + \varepsilon N(0,\eta^2) \qquad (4)$$

where the parameter ε gives the proportion of outliers in the data set while $\eta > 1$ is related to the size of outliers. Letting $\overline{x_n}$ denote the sample mean, two competing measures of scatter that Huber considered are the mean absolute deviation, $d_n = \sum \left| x_t - \overline{x_n} \right| / n$ and the mean square deviation $s_n = \sqrt{\sum \left(x_t - \overline{x_n} \right)^2 / n}$. A standard result of robustness analysis is that for given η and $\varepsilon > 0$, d_n is asymptotically more efficient than s_n. Let us consider now that the sample size is fixed, say n=100, but that ε and η are random quantities. We further develop the previous analysis by focusing on the sensitivity of d_n and s_n to the proportion and to the size of outliers. Describing the uncertainties around ε and η as uniformly distributed over (0,.2) and (2,4) respectively, it is then possible to implement a sensitivity analysis using the method of Sobol'. Mainly, this consists in
sampling $\varepsilon^{(j)} \sim U_{(0,.2)}$ and $\eta^{(j)} \sim U_{(2,4)}$ for $j = 1,\ldots,R$ and $R = 1000$,
simulating for every couple $(\varepsilon^{(j)}, \eta^{(j)})$ n=100 observations from (4), and
computing $d_n^{(j)}, s_n^{(j)}$.
On that basis Sobol's first order indices are:

Statistics considered	Sobol's first-order index for the size of outliers η	Sobol's first-order index for the proportion of outliers ε
d_n	.10	.39
s_n	.16	.34

Hence d_n and s_n are more sensitive to the proportion of outliers than to their sizes. Also s_n is slightly more sensitive to the size of outliers than d_n. For both statistics d_n and s_n there is an important interaction term between η and ε as the sum of the first order terms is much smaller that one.

Conclusion. Two important differences between robustness and sensitivity analysis can be seen: first, the perception of uncertainty is more general in SA since any source can be handled while robustness concentrates on departures from a distribution assumed for the data. Second, SA is able to match output variance with uncertainty sources while robustness focuses on output variance, not on the variance components. With respect to this last point, both can be seen as complementary since robustness give message about overall reaction to uncertainties while SA provides information about the structure of the reaction.

3 FIRST TEST CASE

3.1 Introduction

In the context of environmental modelling, model factors can be affected by large uncertainties. The extended FAST has been performed to identify those model factors which essentially drive the uncertainty on the decision to be made. SA can help the modeller in improving the process of building up highly aggregated information (e.g. pressure indices), i.e. in reducing the overall output uncertainty below reasonable levels.

The extended FAST will be used to apportion the output uncertainty to different subgroups of factors. Factors can be logically regrouped in different fashion, e.g. controllable versus uncontrollable sources of uncertainty, or uncertain data versus uncertain weights, etc.

"Controllable" sources of uncertainty are, for example, the errors due to lack of harmonization in official statistics. "Uncontrollable" refers e.g. to stochastic processes, or lack of consensus in the scientific community, or uncertainty e.g. on temporally distant impacts of climate change that would have to be included in the model.

The model that has been considered in this project contains 196 uncertain factors. They have been categorised into ten groups, and the total sensitivity indices have been estimated for each group. The groups of uncertain factors are hereafter. Associated with each factor is a probability distribution function and range, which are shown in Table 3.1.

This project explores the "sensitive points" of the chain *data collection – data processing – calculation of indicators – calculation of indices – usage of indices for decision-making*, trying to give recommendations on where improvements to data quality and/or methodologies would yield the maximum added value for decision-makers using environmental indicators.

A problem of environmental management is addressed. It has been assumed that two exclusive options for the disposal of solid waste are available to the policy-maker, who must choose the "right" option between incineration and landfill. The model employed to assist the policy-maker supplies a pressure-on-decision index (PI) for each policy option. The index is proportional to the overall hazard to the environment of the corresponding option. The model output is defined as a suitable combination of the pressure-on-decision indices for the two options

$$Y = log\left(PI_{incineration} \Big/ PI_{landfill} \right)$$

so that positive values for the output indicate that landfill is the preferred option, and vice-versa.

3.2 Model structure and uncertainties

In the first stage, the waste management model evaluates a set of environmental pressure indicators, by aggregating atmospheric emissions through weighting coefficients (the intra-weights). These weights measure the harmfulness of a given pollutant to a specific environmental theme (e.g., greenhouse effect). Pressure indicators are defined on a set of selected environmental themes (see Figure 3.1 for the list of themes suggested by Eurostat).

Then, aggregating pressure indicators by means of inter-weights yields a PI. Inter-weights are weighting coefficients attached to the environmental themes, representing their relative importance in terms of the belief of a given stakeholder.

The model embodies a chain of uncertainties.

Sources of uncertainty in data collection include roughly estimated statistics, technical coefficients (either evaluated by expert judgement, estimated via in-situ measurements, or derived from the scientific literature), and the choice of aggregation levels for gathering data, as these are available at different spatial resolutions (i.e. Nomenclature of Territorial Units for Statistics, NUTS).

Uncertainties in indicator building include the selection of the environmental themes for inclusion in the aggregation model, and the choice among different intra-weights for a specific environmental theme.

In pressure indices building, uncertainties affect either policy/sustainable targets (Adriaanse (1993)) or stakeholders' preferences (Puolamaa (1996)).

Groups and Cardinality	*Definition*	*p.d.f.*	*Range (values)*
TU (1)	Territorial Unit: two levels of aggregation for collecting data	Discrete uniform	0 ≡ national level 1 ≡ regional level
DATA (176)	Activity Rates (120) and Emission factors (37)	Uniform	Nominal value (NV) from Corinair database Range: 2 QF × NV; Quality factors (QF): A=5%, B=10%, C=20%, D=50%, E=90%
	National Emissions (19)	Normal	$\sigma = 0.1 \times$ nominal value
E/F (1)	Approach for evaluating indicators: Eurostat or Finnish set	Discrete Uniform	0 ≡ Eurostat list (Figure 3.1) 1 ≡ Finnish list
GWP (1)	Weights for Greenhouse indicator (in the Finnish set): 3 time-horizons	Discrete Uniform	0 ≡ GWP at 20 y 1 ≡ GWP at 100 y 2 ≡ GWP at 500 y
W_E (11)	Weights for Eurostat indicators	Uniform	Nominal value from Scientific Advisory Groups Range = ± 20% of nominal value
EEC (1)	Approach for evaluating environmental concerns	Discrete Uniform	0 ≡ Target Values (Adriaanse (1993)) 1 ≡ Expert Judgement (Puolamaa (1996))
ADR (1)	Type of target values according to Adriaanse (1993)	Discrete Uniform	0 ≡ policy targets 1 ≡ sustainable targets
SUS (2)	Sustainable target for Ozone Depletion	Uniform	Range (0.01;0.10)
	Sustainable target for Dispersion of Toxics and Eco-toxicological Effect	Uniform	Range (0.01;0.02)
F/G (1)	Type of expert judgement: Finnish or German stakeholders	Discrete Uniform	0 ≡ German stakeholders (RSD survey, see Figure 3.2 and Table 3.2) 1 ≡ Finnish stakeholders (see Table 3.3)
STH (1)	Preferences of Finnish Stakeholders	Discrete Uniform	i ≡ stakeholder i i =1,8;

Table 3.1: The ten groups of uncertain factors: the total number of factors in the analysis is 196.

Furthermore, different modules can be employed in some parts of the model, e.g. policy targets for a given year are alternatives to sustainable targets as a design for inter-weights, or Eurostat indicators rather than Finnish indicators), giving rise to 'structural' uncertainties[1].

3.3 Data availability

Atmospheric emissions (e.g., Kg of CO_2 emitted per year), emission factors (e.g., Kg of CO_2 emitted per Tonn of waste incinerated) and production rates (e.g., Kg of municipal waste incinerated) are collected from the CorinAir data base, which is maintained by the European Environment Agency (EEA (1996)). All these emission data are available in a spatially disaggregated form, i.e. either at national level (NUTS0), or at a more detailed level (NUTS1).
CorinAir data include nominal production rates and emission factors for a given type of waste, according to possible destinations (SNAP codes, i.e. Selected Nomenclature for Air Pollution).
The pollutants considered in this study are $SO_2, NO_x, NMVOC$ (non-methane volatile organic compounds), $CH_4, CO, CO_2, N_2O, NH_3$, dioxins, furans, plus the heavy metals $As, Cd, Cr, Cu, Hg, Ni, Pb, Se, Zn$. CorinAir files often supply quality factors in a qualitative fashion, i.e. A (for best quality data) to E. When the quality is not reported the value E is chosen. Uncertainty is then added to nominal values on the basis of the corresponding quality factor (see Table 3.1).

3.4 Computing air emissions

Two types of waste, municipal and industrial, are considered in this study. Municipal production rates are obtained by collecting and aggregating incineration of municipal waste, open burning of agricultural waste, landfill and compost production. Industrial production rates are obtained by aggregating incineration of industrial waste and landfill. Given that a vague idea exists about the amount of waste coming from municipalities that is placed in landfill, we assumed that 70% is municipal and the remaining 30% industrial.[2] Emission factors are poor, and sometimes missing. They are often integrated with data suggested in the literature (EEA (1996)). Air emissions are finally evaluated by weighting production rates with the corresponding emission factors.

3.5 Pressure indicators

Two alternative sets of indicators have been used, the set of Finnish indicators proposed in (Puolamaa (1996)), and the list of Eurostat indicators (Eurostat (1999)).
Finnish Indicators
Three Finnish indicators have been employed: the greenhouse indicator, the acidification indicator and the indicator of eco-toxicological effect.
The greenhouse indicator is calculated by considering the effects of carbon dioxide, methane and nitrous oxide emissions (i.e., N_2O and NO_x). The acidification indicator refers to the atmospheric deposition of acidifying compounds, and is derived by taking into account the effect of sulfur dioxide, nitrous oxides (NO_x) and ammonia emissions, which are the most acidifying compounds. The indicator of eco-toxicological effect includes heavy metals in the list above, and organic dibenzodioxines and -furans.
When evaluating the greenhouse indicator, the global warming potentials (GWPs) represent a source of uncertainty because three sets of GWPs have been proposed by the Intergovernmental Panel for Climate Change (Houghton et al (1996)) according to three different time-horizons (see Table 3.1). Given that there is no scientific basis for choosing any one of them, this (subjective) uncertainty has been tackled in the exercise, by considering the three sets of values as occurring with the same probability.
For the acidification indicator the intra-weights for the various compounds are represented by acidification potentials, which are defined on the basis of the quantity of hydrogen ions that would be formed in

[1] We term 'structural' those uncertainties involving the question "was the right model selected?".
[2] Very poor estimates of this fraction are currently available. It is noteworthy that the model has revealed the necessity to have this kind of distinction, which is another source of uncertainty (although not yet quantitatively included here, but planned to be accounted for in the future).

complete atmospheric deposition of the gas in question. The values for the acidification potentials adopted in this study, expressed as kilos of sulfur dioxide equivalents per kilogram of the compound considered, are taken from (Puolamaa (1996)).
The intra-weights necessary to implement the eco-toxicological indicator are the eco-toxicity potentials (ETP). They give a measure of the persistence of toxic substances accumulated in the aquatic or terrestrial ecosystem. Weights for heavy metals have been gathered from (ICI (1997)).

3.6 Eurostat Indicators

In order to define the contents and structure of the Pressure Indices project, Eurostat organized surveys among natural scientists, the so-called Scientific Advisory Groups (SAG) (Jesinghaus (1999)). The SAG ranked the indicators within each environmental theme in order of importance. The Eurostat core list of environmental themes and the six most influential indicators for each theme are displayed in Figure 3.1 from left to right in order of decreasing importance (Eurostat (1999)). The rankings are expressed by means of intra-weights, which constitute the group W_E of uncertain factors (see Table 3.1). The intra-weight for a given indicator is defined as the percentage of SAG experts who selected that indicator for the core list.
Four themes have been selected because of their relevance to the study: Air Pollution, Climate Change, Ozone Layer Depletion and Dispersion of Toxic Substances. The theme "Waste" itself is, of course, also relevant for the study. However, the SAG of waste experts has expressed its subjective preference, attributing a higher score to landfill than to incineration (Figure 3.1). Given that it would be a double-counting to include the implicit valuation of the waste experts at this level, we have decided not to include the theme Waste in the analysis, but to use it as a reference for interpreting the results of the more detailed analysis performed here.
In the theme Air Pollution the emissions of NO_x, NMVOC and SO_2 are included in the model. Particulate emissions have not been included for lack of data.
Waste management activities affect the theme Climate Change through all five indicators; however, emissions of chlorofluorocarbons (CFCs) are not included because of lack of data on emission factors.
In the theme Ozone Layer Depletion, emissions of carbon dioxide and of nitrogen oxides (NO_x) are included. Emissions of halons, CFCs and HCFCs are omitted for lack of data on emission factors.
In the theme Dispersion of Toxic Substances, emissions of persistent organic pollutants (dioxines and furans), to air (for incineration) and of heavy metals (to air for incineration, to groundwater for landfill) are considered.

3.7 Indicators of total burden

Indicators at national level describe the total burden of human activities in a given country. These indicators are needed both for computing target values (section 3.8.1) and for indicator normalisation. In the present experiment, depending on the set of indicators that is used (either Finnish or Eurostat) in a specific run of the model, the corresponding set of indicators at national level has to be evaluated. Firstly, atmospheric emissions from all the economic sectors and for all the pollutants of interest are collected. Emissions of $SO_2, NO_x, NMVOC, CH_4, CO, CO_2, N_2O, NH_3$ and heavy metals are gathered from Corinair at NUTS0, while dioxines and furans national emissions are provided through a survey conducted by the Landesumweltamt NRW (*German* State Environmental Agency (1997)). Finally, national emissions are aggregated to indicators of total burden as illustrated in section 3.5.

3.8 Pressure-to-decision indices

The inter-weights used to aggregate indicators to indices must be representative of the perceived impact of the problem covered by the themes, and should include both objective and subjective elements. An objective approach should be able to quantify the distance of the actual situation from a policy target[3], whereas a subjective approach should take into account experts' and public opinion. Here we present two alternative approaches: the former is based on objective criteria (section 3.8.1), the latter encompasses

[3] "Objective" in the sense of "observable", which does not necessarily mean that a political decision for a specific target is scientifically and/or economically sound.

subjective aspects (section 3.8.2). The triggering factor [ADR] has been introduced to select between the two approaches, which have the same probability of occurring in each model execution.

3.8.1 Target values

In the objective approach a target for each theme indicator is set, and the inter-weight is defined as the reciprocal of the target value (Adriaanse (1993)).
Targets can be formulated by relating either to policy targets, which are chosen for a certain reference year, or to sustainability levels, representing the targets proposed to reach an environmental sustainable development.

3.8.2 Expert judgement

The subjective approach is based on expert judgement (Puolamaa (1996)). The pressure indicators are first normalised using the total national burdens and then fed into the inter-weighting stage. Two sets of inter-weights are used in this study. A trigger factor [F/G] has been set up to randomly select one of the two sets of inter-weights.
In the first set a group of Finnish stakeholders from eight different domains (Table 3.2) were selected and interviewed (Puolamaa (1996)). The Analytical Hierarchy Process (AHP) performed in (Puolamaa (1996)) provides the weighting of the environmental problems by stakeholders.

	Greenhouse Climate Change	**Eco-toxicologic. Dispers. Toxics**	**--- Ozone layer Depl.**	**Acidification Air Pollution**
Agriculture	15.6	11	14.1	6.5
Manufacture	14.2	10.6	14.2	7.3
Env. Journ.	15.5	10.7	17.1	4.6
Env. NGOs	11.8	9.3	14	3.3
Administration	20	8.1	14.1	5.2
Traffic	13.8	11.2	8.7	6.7
Env. Scientists	22.3	8.7	10.5	5.1
Politicians	9.9	10.4	16.5	5.5
All groups	**15.4**	**10**	**13.7**	**5.5**

Table 3.2. Result of the Analytical Hierarchy Process. Weighting of environmental problems by stakeholders.

A subjective source of uncertainty is due to the fact that different weights, associated with the various domains (first column of Table 3.2), are available for each environmental indicator. In a typical deterministic analysis the average inter-weight (last row in Table 3.2) would be employed. In the present study we allow each domain to occur with the same probability in a given model execution, thus taking this source into account (factor [STH]).
The second set of inter-weights is obtained from a survey performed by the Research Unit for Societal Development, University of Mannheim in 1990, where a group of German experts from different sectors (Figure 3.2) were asked to provide a weighting coefficient for various themes (Table 3.3).

Finnish ind → Eurostat ind →	**Greenhouse (Climate Change)**	**Eco-toxicol. (Dispers. toxics)**	**--- (Ozone layer Depl)**	**Acidification (Air Pollution)**
Average Weighting factors	**0.166**	**0.12**	**0.11**	**0.12**

Table 3.3. Inter-weights obtained from the survey conducted by the Research Unit for Societal Development, University of Mannheim in 1990.

3.9 Results and discussion

Three cases are illustrated in this section: they relate to Austria in the year 1994.
The first exercise is carried out on a preliminary version of the model where the indicators of eco-toxicological effect (ET) and of Dispersion of Toxics (DT) are not taken into account, and the factors [ADR, SUS and F/G], are not sampled. The bimodal histogram displayed in Figure 3.3 represents the outcome of the uncertainty analysis. The left-hand region, where incineration is preferred, comprises approximately 40% of the total area, and this shows that the model does not allow a ranking of the two options. The pie chart in Figure 3.4 confirms that the choice of the indicator set has an overwhelming influence and reveals that the scientific community must work to find a consensus on the proper set. Here sensitivity analysis does not help to decide whether one should use the Finnish or the Eurostat indicators, but about the relative importance of the various types of uncertainty, including that arising from the choice of indicators.
In the second experiment, sensitivity analysis has been performed for the two cases separately, under the hypothesis that a consensus among the experts has been already established on the use of a given set of indicators.
Figure 3.5.a illustrates the results when employing the Finnish set, whereas Figure 3.5.b illustrates the use of the Eurostat set. In Figure 3.5.a, data account for 45% of the total output variance V (this is expected, due to the poor quality of waste data). The GWP factor shares 40% of V, thus indicating that more effort is required from the IPCC Panel of experts in order to converge towards a more stable set of GWP values. The remaining 15% includes all the other factors.
In summary, "data uncertainty" (the factor [Data]) is about as important as “model uncertainty” (all the other factors), i.e. the uncertainty introduced by the analyst. This situation could lead to criticism and would not be acceptable in most practical applications. If a reduction in output uncertainty were required in order to better support an eventual decision, the analyst would know how to prioritise his/her data acquisition effort.
Figure 3.5.b shows that the impact of “data uncertainty” on the PI is very large (82%), underlining high model relevance and quality, the model being weakly influenced by subjective uncertainties.
In the third exercise the eco-toxicological and dispersion indicators have been included in the model, and all groups of factors have been considered. Here a consensus on the policy has been reached (landfill is preferable) although the uncertainty in the resulting output is quite large (histogram in Figure 3.6).
In summary, the results reported above show the importance of including dioxins and furans emissions in the model. The design of data collection should try to strike a balance between accuracy and complete coverage. A few “precise” indicators may please the eye of the statistician, but a more complete “optimally inaccurate”[4] set better fits the requirements of decision-making processes and can yield more reliable support for the ultimate decision.
The sensitivity analysis for the second exercise (Figure3.7) shows that the total impact of “data uncertainty” on the output index is low (37%), the largest fraction being due to “subjective” uncertainties. Addressing further investigation on E/F and GWP (15% each) will enhance model accuracy.
These results strengthen the role of SA as an adequate tool for helping the analyst to test and improve the quality of the model and the transparency of the decisions. A reduction of the model output uncertainty below reasonable levels increases the likelihood for a certain environmental policy to attain the suitable scientific added value that may support its implementation.
Other indicators could be further included in the model: particulate matter emissions, which have a negative impact on incineration, are still not considered. However, this study is purely methodological. The focus is indeed on the novelty of the approaches used to tackle the problem of uncertainties in environmental policy. In standard approaches single (nominal) values are given to uncertain model factors, resulting in outputs with no confidence levels attached. Such a situation can hardly be relevant to decision-making.

[4] The notion of “optimally inaccurate indicators” was developed by John O’Connor, formerly World Bank (personal communication).

The model used in this work could be employed to examine the combination of other policy options. For instance, it could be used to study what happens if municipal waste is placed in landfill and industrial/hospital wastes incinerated, or what are the effects of incinerating 40% of the total production and the rest placed in landfill, whatever the composition of waste.
A generalisation of the model could allow the investigation of other interesting policy options as, for instance, the quantification of the consequences of banning the open burning of agricultural waste in rural regions, or the effect of using modern incineration plants for industrial waste instead of old ones without particle abatement systems. A more complex generalisation could involve the inclusion of socio-economic aspects of environmental change, in the perspective of making policies for sustainable development.

4 SECOND TEST CASE: APPLICATION TO SIGNAL EXTRACTION IN ECONOMIC TIME SERIES

4.1 Introduction

A common practice in official statistics and applied macroeconomics is to focus the analysis on some unobserved movements like trend and seasonally adjusted series. Several methods are available for trend extraction and seasonal adjustment. With model-based methods (see for example Bell and Hillmer, 1984), the estimation of the unobserved components is carried out after selection of a model and maximum likelihood (ML) estimation of the parameters. When several models describe the data dynamics in a statistically correct way, an information criterion is typically used to select a single specification. Given the data, the model chosen and the parameter estimates, the unobserved components are computed by signal extraction either via Wiener-Kolmogorov or Kalman filtering.

Such unobserved component estimators can be seen as conditional on a model specification and on the estimated parameters values. In practice however, time series models are only description of reality, so different models may describe the data equally well. A first problem for an official statistician who needs to produce trends and seasonally adjusted series is thus: within the class of the acceptable models, how sensitive are the unobserved component estimators to the model choice? Further, for a given model specification, the collection of new observations may slightly modify the ML parameter estimates. In consequence, the unobserved component estimators would present some instability due to the model coefficients update. Such instability is highly undesired by practitioners. A second problem is thus: for a given model, how sensitive are the unobserved component estimators to changes in model parameters?

In this test case, our intention was:

- to assess the uncertainty of the component estimators due to the uncertainty of model specification and of parameters, and
- to apply sensitivity analysis techniques to isolate the most influential sources of uncertainty for the characterisation of the trend, seasonal and short-term movements in time series.

This second objective was studied through the implementation of sensitivity analysis techniques. Some related works can be found in Carlin and Dempster (1989) mainly about parameter uncertainty in fractionally integrated Auto-Regressive Moving Average (ARMA) models, and in Bell and Otto (1993) about the joint effect of parameter and some type of data uncertainties. Parameter and model uncertainty have been considered in a forecasting context by Draper (1995). Our development can be distinguished from these works first as it addressed the problem of model and parameter uncertainty in the unobserved component models framework, and second as it applies sensitivity analysis techniques.

It is worth mentioning that other sources of uncertainty exist in the model-based signal extraction framework. Drawing an exhaustive list of uncertainty sources, Bell (1989) also included: (1) sampling and non sampling errors in data; (2) decomposition uncertainty; (3) signal extraction error. We choose to ignore them on the basis of the following considerations. First, we suppose that the series at hand does not result from a survey, so it will be free of sampling errors. Second, the effect of decomposition uncertainty on the unobserved component estimators can be assessed analytically (see Maravall and Planas, 1999). We shall use the same identification assumption throughout the analysis, namely the so-called *canonical* decomposition (see Pierce, 1978). Third, we view signal extraction error as a consequence of the other uncertainties. This is so because we believe that it is completely determined by model, parameters, decomposition choice, and the sampling/non sampling data error structure if present.

Focusing on trend, seasonal and irregular movements in an economic time series we developed an in-depth analysis of the effects of parameter uncertainty. Next, for a given range of models, we analysed how sensitivity to model specification can be evaluated. The analysis put together three main tools:

- the model-based signal extraction techniques (see Burman 1980);

- a Bayesian framework for characterising the distribution of model parameters and for assigning probability to time series models (see Kass and Raftery, 1995);
- sensitivity analysis techniques, and we mainly relied on Saltelli et al. (1993).

4.2 General procedure

The implementation of uncertainty and sensitivity analysis tools in the time series framework needs three main ingredients:

- an observed series,
- a parametric model which provides the analyst with a description of the uncertainties, and
- a functional of interest.

By parametric model, it is intended a set of parameters associated to a model specification. In our study, we focused on regression with time series errors, with the stochastic part decomposed in unobserved movements such as trend, seasonal and irregular components. Information on parameter uncertainty is obtained by combining via Bayes theorem the likelihood function with a prior density for the parameters. This yields a posterior distribution which characterises the parameters uncertainty (see Box and Tiao, 1973). For considering model uncertainty, a set of parametric models is needed so that a model averaging can be performed (see Draper, 1995). A probability is then assigned to every model in the set of model considered using the Bayes factor (see Kass and Raftery, 1995). Finally, sensitivity analysis requires to state explicitly the objective of the analysis; we designed a functional especially for addressing the issues of model assessment with respect to unobserved movements and of robustness of unobserved component estimators.

With these elements available, the effect of uncertainties on the functional of interest could be characterised. The overall variance was a first summary. We could see that simple plots of the functional values with respect to each parameter bring a relevant information, as discussed in Young (1999). Sensitivity analysis enables to quantify how much of the overall variance is related to every uncertainty sources. We made use of importance measures (see Saltelli et al., 1993). As discussed in the previous reference, importance measures can be seen as first-order Sobol's sensitivity indexes. These points have been discussed using an application to an actual economic series. The example that we present in this report illustrates the benefit obtained from considering sensitivity analysis techniques in modelling economic time series: mainly,

- sensitivity analysis gives some information about how accurately a model or a set of models defines a target of interest.
- it brings an insight into model assessment
- it enables an analyst to isolate the model with best properties with respect to the objective of the study.
- sensitivity analysis can be seen as another way of looking to model parsimony, complementary to information criteria like the Bayesian information criterion (BIC) (see Hannan, 1980).

4.3 Technical framework

We considered the time series regression framework (see Fuller, 1996) where the observations Y_t, $t = 1,...,T$, are represented as:

$$Y_t = X_t' \beta + y_t \qquad (6)$$

where X_t is a vector of *r* deterministic regressors, possibly describing outliers or calendar effects like trading days or length of month. An ARIMA model was supposed to characterise the stochastic term y_t according to

$$\delta(B)\phi(B)y_t = \theta(B)a_t \qquad (7)$$

where a_t is a white noise with variance V_a, $\delta(B)$ is a differencing operator, and $\phi(B),\theta(B)$ are finite polynomials in the lag operator B of respective order p, q, which verify the satisfy the stationarity and invertibility conditions. The set of model parameters will be denoted α, $\alpha=(\phi_1,...,\phi_p,\theta_1,...,\theta_q,\beta_1,...,\beta_r)$. Once a set of regressors is chosen, the specification and estimation of model (6,7) can be performed using standard time series techniques (see for example Brockwell and Davis, 1991). We shall denote the likelihood function $p(y\mid M,\alpha,V_a)$, where *M* underlines that it depends on the model specification. Formally, letting y_D,Y_D,X_D denote the series obtained by the differencing operations $\delta(B)y_t,\delta(B)Y_t,\delta(B)X_t$ and $\Sigma_D=V[y_D]$, then

$$p(Y\mid M,\alpha,V_a)=p(Y_D\mid M,\alpha,V_a)$$
$$\propto \left|\Sigma_D\right|^{-\frac{1}{2}} exp[-\frac{1}{2}(Y_D-X_D'\beta)\Sigma_D^{-1}(Y_D-X_D'\beta)] \qquad (8)$$

The ML estimators will be denoted $\hat{\alpha}\equiv\hat{\alpha}(M)$. In a second stage, the stochastic term is supposed to be made up of orthogonal unobserved components according to

$$y_t=p_t+s_t+u_t \qquad (9)$$

where p_t is a trend, s_t a seasonal component, and u_t an white-noise irregular component with variance V_u. The stochastic components p_t and s_t are also described by ARIMA models as

$$\phi_p(B)p_t=\theta_p(B)a_{p_t}$$
$$\phi_s(B)s_t=\theta_s(B)a_{st} \qquad (10)$$

where $\phi_\ell(B),\theta_\ell(B)$, $\ell=p,s$ denote finite polynomials in *B*, with all roots on or outside the unit circle. The variables a_{pt} and a_{st} are independent white noise with variances V_p and V_s, respectively. The polynomials $\phi_p(B)$ and $\phi_s(B)$ are prime, while the moving average polynomials $\theta_p(B)$ and $\theta_s(B)$ share no unit roots in common.

We identify a unique decomposition (9,10) by imposing the canonical requirement on the models for the trend and seasonal components (see Hillmer and Tiao, 1982). The unobserved quantities can then be estimated by signal extraction techniques as

$$\hat{p}_t=E[p_t\mid Y] \qquad (11)$$

and similarly for the other components (see Burman, 1980). The allocation of the regressors in (6) to the unobserved components is performed by matching the regressor-type with the properties of the components; for example, calendar effects are assigned to seasonal stochastic component, while a shift in level would be assigned to the trend component. Model (6-10) is routinely specified, estimated and decomposed using the software TRAMO-SEATS (see Gomez and Maravall, 1996).

Making explicit the conditioning on model specification and parameter estimates, equation (11) can be written as

$$\hat{p}_t = E[\, p_t \,|\, Y, M, \hat{\alpha}\,] \tag{12}$$

Our aim is to investigate the sensitivity of the estimates to the model specification and to the model parameter estimates. We shall focus on mean squares deviations of the extracted component around the estimator related to ML parameter values:

$$f_p(\alpha, M) = \frac{1}{T}\sum_{t=1}^{T}(E[\, p_t \,|\, Y, M, \alpha\,] - E[\, p_t \,|\, Y, M, \hat{\alpha}\,])^2 \tag{13}$$

The mean square deviations around seasonal and irregular components estimators related to the ML model parameter values are similarly measured using $f_s(\alpha, M)$ and $f_u(\alpha, M)$.

We now apply a Bayesian framework to the treatment of uncertainties. A posterior distribution characterising the uncertainty around the model parameters is obtained by combining the likelihood function with a prior probability:

$$p(\alpha, V_a \,|\, Y, M) \propto p(Y \,|\, M, \alpha, V_a)\, p(\alpha, V_a \,|\, M) \tag{14}$$

Following Bell and Otto (1993), we use a non-informative prior on α, V_a over the parameter space where the ARMA parameters lie inside the stationarity and invertibility regions. Integrating (14) over V_a yields the marginal posterior distribution $p(\alpha \,|\, Y, M)$. We use an asymptotic approximation given by the normal distribution with mean $\hat{\alpha}$ and variance $\hat{\Sigma}$, the negative inverse Hessian of the log-likelihood evaluated at the ML parameters (see Box and Jenkins, 1970, p.252-258). This distribution is subject to a truncation due to the stationarity and invertibility conditions. More elaborate techniques for simulating from the exact posterior can be implemented; for the time being, we rest with this approximation in order to concentrate the discussion on the application of sensitivity analysis.

The problem of assigning a posterior probability to a model, $p(M \,|\, Y)$, has recently received much attention in the statistical literature; see Kass and Raftery (1995) for an exhaustive survey. In this application, use is made of the Bayes factor. In short, if $M_1, \ldots, M_K$ are *K* model considered,

$$p(M_k \,|\, Y) = \frac{p(Y \,|\, M_k)\, p(M_k)}{\sum_{i=1}^{K} p(Y \,|\, M_i)\, p(M_i)} \tag{15}$$

where $p(M_i)$ is a prior probability on model *i*, and

$$p(Y \,|\, M_k) = \int p(Y \,|\, \alpha, V_a, M_k)\, p(\alpha, V_a \,|\, M_k)\, d\alpha\, dV_a\,, \tag{16}$$

$p(Y \,|\, \alpha, V_a, M_k)$ being the likelihood function. In the regression with time series errors framework, only models whose residuals pass all usual diagnostic checks can be considered. This requirement restraint considerably the range of plausible models, so in practice we are led to consider a limited number of possibilities. None of these models are true, since they only describe in a satisfying way the series properties. In that context, it is difficult to anticipate that for a given series, a model will be more likely. We shall thus be 'neutral' and impose the same prior on every model, so $p(M_k) = p(M_i)$, $k \neq i$. Expression (15) simplifies to

$$p(M_k \mid Y) = \frac{p(Y \mid M_k)}{\sum_{i=1}^{K} p(Y \mid M_i)} \tag{17}$$

Several approaches are available for evaluating the integral (16). Following Kass and Raftery (1995) and Draper (1995), we used a Laplace approximation evaluated around the maximum likelihood estimators and we considered the large sample approximation $ln\left|\hat{\Sigma}\right| \approx -n_k \, ln\, T$, where n_k is the number of parameter associated to model M_k . This leads to

$$ln\ p(Y \mid M_k) = \frac{n_k}{2} ln(2\pi) - \frac{1}{2} BIC(M_k) + O(1) \tag{18}$$

where $BIC(M_k)$ is the Bayesian information criterion associated with model M_k (see Hannan, 1980). So the BIC of every fitted model provides a simple way through (18) to assign a probability to every model.

4.4 Application to an economic time series

Our test case is the monthly series of Italian Production of Soft Drinks (IPSD), which is made up of 107 observations between January 1985 and November 1993. Three models describing correctly the data are detailed below, with the ML parameter estimates and standard deviations in parenthesis.

Model 1: $(2,1,1)(0,1,1)_{12}$, $\alpha = (\phi_1, \phi_2, \theta_1, \theta_{12})$

$$(1 + \phi B_1 + \phi_2 B^2)\Delta\Delta_{12} y_t = (1 + \theta_1 B)(1 + \theta_{12} B^{12}) a_t$$

$\hat{\phi}_1 = -.079$ (.130) $\hat{\phi}_2 = .297$ (.116) $\hat{\theta}_1 = -.733$ (.106) $\hat{\theta}_{12} = -.606$ (.117)

Model 2: $(0,1,1)(0,1,1)_{12}$, plus trading days and length of month regressors; $\alpha = (\beta_1, \beta_2, \theta_1, \theta_{12})$

$$Y_t = \beta_t x_t + \beta_t x_t + y_t$$
$$\Delta\Delta_{12} y_t = (1 + \theta_1 B)(1 + \theta_{12} B^{12}) a_t$$

$\hat{\beta}_1 = .004$ (.002) $\hat{\beta}_2 = .059$ (.044) $\hat{\theta}_1 = -.763$ (.073) $\hat{\theta}_{12} = -.658$ (.120)

Model 3: $(3,1,0)(0,1,1)_{12}$, $\alpha = (\phi_1, \phi_2, \phi_3, \theta_{12})$

$$(1 + \phi B_1 + \phi_2 B^2 + \phi_3 B^3)\Delta\Delta_{12} y_t = (1 + \theta_{12} B^{12}) a_t$$

$\hat{\phi}_1 = .506$ (.105) $\hat{\phi}_2 = .527$ (.010) $\hat{\phi}_3 = .202$ (.107) $\hat{\theta}_{12} = -.733$ (.134)

We can see that the three models are not nested. Table 4.1 reports several diagnostic checks on the residuals. The diagnostics displayed are namely the Ljung-Box statistics on the first 24 auto-correlations $Q_{24}(a_t)$, the Box-Pierce statistics on the first two seasonal lags, skewness (*S*) and kurtosis (*K*) aggregated to form the Bowman-Shenton test for normality *N* (see Bowman and Shenton, 1975). The BIC associated to every model is also reported.

	Model 1	Model 2	Model 1
$\sqrt{\hat{V}_a}$	.0695	.0697	.0706
$Q_{24}(a_t)$	28.16	29.61	26.18
$Q_{s2}(a_t)$	.21	.25	.92
1 N= $S^2 + K^2$	1.49	2.25	2.22
BIC	-213.83	-212.22	-207.51

Table 4.1. Diagnostics for models 1,2 and 3.

The three models give a satisfactory description of the IPSD series. In front of these three ways of describing the data dynamics, the obvious problem of a practitioner who wishes to produce a trend or a seasonally adjusted series is: which model to choose? Focusing on the BIC would favour model 1 against model 2 and model 2 against model 3. But BIC does not tell us much about how well these models characterise the structure of the series. Yet the three specifications imply some differences in the modelling of the components. In models 1 and 3, the AR polynomials have roots at frequencies close to the seasonal harmonics, so the AR polynomial is associated with the seasonal movements. So models 1 and 3 imply a decomposition of the type:

$$\Delta^2 p_t = \theta_p(B) a_{pt}$$
$$\phi(B)(1 + B + ... + B^{11}) s_t = \theta_s(B) a_{st} \qquad (19)$$

In model 2, the regressors involved represent the trading days rhythm (constructed as number of working days minus number of Saturdays and Sunday times 5/2) and the length of months. When decomposing the series under model 2, these regressors will be assigned to the seasonal components, and the decomposition is:

$$\Delta^2 p_t = \theta_p(B) a_{pt}$$
$$s_t = s_t' + \beta_1 x_{1t} + \beta_2 x_{2t}$$
$$(1 + B + ... + B^{11}) s'_t = \theta_s(B) a_{st} \qquad (20)$$

The irregular term remains a white noise in the three cases. The first question that we consider is: given the parameter uncertainty related to every model, how well every model characterises the series structure?

We simulate *R* values for the parameters α of each model according to $\alpha \sim N(\hat{\alpha}, \hat{\Sigma})$. The estimated covariance matrix $\hat{\Sigma}$ is given by the software used, TRAMO. So, as long as the normal approximation is deemed appropriate, no difficulties are involved in this stage. For every simulated value $\alpha^{(j)}$, *j=1,…,R*, the observed series model is decomposed into unobserved components, which are estimated by signal extraction. It is then straightforward to compute the sample mean square deviations (Equation 13) for every unobserved component estimator, $f_p(\alpha^{(j)})$, $f_s(\alpha^{(j)})$, $f_u(\alpha^{(j)})$. Table 4.2 below reports the mean and variance over 1000 replications of the sample mean square deviations. It can be seen that model 2 is the model which entails the less uncertainty around the unobserved movements: the sample deviations around the estimators derived from ML parameters values are much lower in mean and these deviations have a lower variance with respect to the other 2 models.

	Model 1	Model 2	Model 3
$\overline{f_p}(\alpha)$	.015	.003	.022
$Var[f_p(\alpha)]$ $(\times 10^{-3})$	.757	.014	1.032
$\overline{f_s}(\alpha)$	.050	.017	.150
$Var[f_s(\alpha)]$ $(\times 10^{-3})$	4.900	.208	52.156
$\overline{f_u}(\alpha)$	.041	.016	.138
$Var[f_u(\alpha)]$ $(\times 10^{-3})$	3.332	.146	43.882

Table 4.2, uncertainty analysis for models 1,2 and 3

We now move to sensitivity analysis, and to the model assessment. How much of the deviations around $\hat{p}_t$, $\hat{s}_t$, $\hat{u}_t$ can be attributed to the different parameters ? The *importance measure* discussed in Section 2 enables us to answer that question. Let α be written as $\alpha = (\alpha_i, \alpha_{-i})$, where α_i is a single parameter and α_{-i} embodies all the other elements. Making explicit the conditioning arguments related to our context, the importance measure can be written:

$$IM_{\alpha_i}(f_\ell(\alpha)) = \frac{VE[f_\ell(\alpha)|Y, M_k, \alpha_i]}{V[f_\ell(\alpha)|Y, M_k]} \tag{21}$$

where $\ell = p, s, u$. For an important contribution of the parameter α_i in characterising a component, a low variance of $f_\ell(\alpha)$ when α_i is fixed is expected. If we consider the uncertainty range for α_i, we want a low mean value $E_{\alpha_i} V[f_\ell(\alpha)|\alpha_i]$. Standard results about variance decomposition give

$$E_{\alpha_i} V[f_\ell(\alpha)|\alpha_i] = 1 - IM_{\alpha_i} \tag{22}$$

Hence parameters which contribute mostly to the description of a pattern are associated with high importance measures. Conversely, low importance measures point to irrelevant parameters with respect to the specific pattern. Tables 4.3.a-4.3.c below displays the results for the 3 models fitted to the series ITPSD.

Model 1

IM	$f_p(\alpha)$	$f_s(\alpha)$	$f_u(\alpha)$
ϕ_1	.085	.052	.018
ϕ_2	.042	.068	.029
θ_1	.669	.163	.096
θ_{12}	.040	.822	.910

Table 4.3.a

Model 2

IM	$f_p(\alpha)$	$f_s(\alpha)$	$f_u(\alpha)$
β_1	.001	.143	.194
β_2	.021	.275	.256
θ_1	.760	.012	.049
θ_{12}	.256	.610	.463

Table 4.3.b

Model 3

IM	$f_p(\alpha)$	$f_s(\alpha)$	$f_u(\alpha)$
ϕ_1	.006	.004	.005
ϕ_2	.011	.005	.005
ϕ_3	.012	.002	.003
θ_{12}	.915	.993	.995

Table 4.3.c

It can be seen that the AR parameters nearly do not participate to the description of the series long-term movements. This is a consequence of the allocation of the AR polynomial roots to the seasonal component. Regarding the MA parameters, θ_1 mostly drives the extracted trend while θ_{12} is mostly important for seasonal and short-term fluctuations. Short-term movements seem well enough defined in models 1 and 2, but not in model 3 where only θ_{12} is determinant. The external regressors of model 2 determines in an important way the seasonal and short-term movements of the series. This analysis tends to show that if the target of the analysis is the trend, then models 1 and 3 are over-parametrised because of the AR parameters. If the target is the seasonal movements, then model 3 is over parametrised.

In conclusion, the sensitivity analysis techniques point to model 2 as the model which lead to a better definition of the series structure. This result does not agree with the BIC which favours model 1 instead. This result shows that sensitivity analysis complements other classical tools in an interesting way, shedding light about the accuracy of model definition with respect to some targets of interest.

We now evaluate the sensitivity of the unobserved component estimates relatively to the model choice. In order to simplify to expressions we write $f_\ell \equiv f_\ell(\alpha(M))$. For a given set of models $\mathrm{M} = (M_1,\ldots,M_k)$, a variance decomposition leads to (see Draper, 1995):

$$V[f_\ell \mid Y] = V_{\mathrm{M}} E[f_\ell \mid Y, M_k] + E_{\mathrm{M}} V[f_\ell \mid Y, M_k] \qquad (23)$$

In expression (23), the first term of the right hand side gives the variance between models of the mean of the sample deviations, and the second term the mean over the models of the variances due to the parameter uncertainty. For every model, the mean and variance of the sample deviations are available in table 4.2; the evaluation of (23) mainly follows from these computations. Let us denote $\pi_i = p(M_i \mid Y)$ given by (17) and evaluated using (18), and $\overline{f}_{\ell,M_k} = E_\alpha[f_\ell \mid Y, M_k]$. Then, over the different models

$$\overline{f}_\ell = E[f_\ell \mid Y] = \sum_{k=1}^{K} \pi_k \overline{f}_{\ell,M_k}$$

$$V[f_\ell|Y] = \sum_{k=1}^{K} \pi_k \frac{1}{R} \sum_{i=1}^{R} (f_\ell(\alpha^{(i)}(M_k)) - \overline{f}_\ell)^2 = \sum_{k=1}^{K} \pi_k (g_{\ell,M_k} - 2\overline{f}_{l,M_k}\overline{f}_\ell - \overline{f}_\ell{}^2) \quad (24)$$

where $g_{\ell,M_k} = \frac{1}{R}\sum_{i=1}^{R} f_\ell(\alpha^{(i)}(M_k))^2$

The first term of the right hand side of (23) is evaluated as

$$V_M E[f_\ell|Y,M_k] = \sum_{k=1}^{K} \pi_k (\overline{f}_{\ell,M_k} - \overline{f}_\ell)^2 \quad (25)$$

Combining (24) and (25) yields an importance measure for model choice

$$IM_M(f_\ell) = \frac{V_M E[f_\ell|Y,M_k]}{V[f_\ell|Y]} \quad (26)$$

Applying these results to our case study, we get $\pi_1 = .67$, $\pi_2 = .30$, $\pi_3 = .03$: the Bayes factor does not give much weight to model 3. Then (26) yields

$$IM_M(f_p) = .075$$
$$IM_M(f_s) = .152$$
$$IM_M(f_u) = .161$$

For our case study, these results show that only a relatively low proportion of the uncertainty around unobserved components depends on the model used: the uncertainty about parameters has more important consequences than uncertainty around models. Next, the effect of model uncertainty is mostly relevant regarding seasonal and the short-term movements of the series, while the series trend is the most 'robust' component over the three models.

4.5 About the normal approximation

Before concluding, we briefly discuss the use of the normal approximation as an approximation to the posterior distribution. In the Bayesian framework, several techniques are available for simulating from a posterior distribution (see Geweke, 1996, and Gilks et al., 1996). The main problem faced in our case was the number of replications needed to evaluate the importance measure (15), namely RxR. We tried three algorithms: acceptance sampling, importance sampling, and the Metropolis Markov chain algorithm. A rejection rate of 98% made the first unsuited. Notice that in a similar context Bell and Otto (1993) obtained a rejection rate of the same order when implementing the acceptance sampling scheme. Importance sampling (see Geweke, 1996) leads to compute RxR weights, since an importance function is needed for computing the conditional expectation in (15), and another one for computing the variance of that conditional expectation. The computation of many weights considerably lowers the execution time. Further, it does not deliver a sample from the posterior distribution against which the normal approximation could be compared. The Metropolis algorithm (see Gilks et al., 1996) was more convenient for our case. Yet, the need for discarding *B* first simulations actually leads to a total of R*x*B discarded steps, which becomes more costly. We used that approach to compare the normal distribution with a sample of 2000 α simulated from the exact posterior. Although not very accurate in the tails of the distribution for the regular moving average parameter, it could be seen that the normal distribution was a reasonable approximation if emphasis is put on computing time.

4.6 Conclusion

We believe that sensitivity analysis has a place within the vast battery of tools available in time series analysis. Model specification, estimation and diagnostic checking constitutes the classical steps for handling time series, which are broadly known as Box and Jenkins (1970) model building methodology. Although there have been many developments in time series analysis, for example around the model specification procedure with the structural modelling (see Harvey 1989), also in the range of specifications available (see for example Tong, 1990), that overall approach remains an unquestionable paradigm. Its application usually results in a model or a set of models which describes the series of interest in a statistically correct way. The implementation of uncertainty and sensitivity analysis comes in a second stage, when the statistical validity of the model considered has been checked and when the target of the analysis has been explicitly specified.

Uncertainty and sensitivity analysis provides then a complementary information about how tightly the objective of the analysis is described given the uncertainties faced. The contribution of every parameter of a given model to the definition of the target can be assessed, and which parameters have the most significant effect on the objective of the analysis can be determined. For a good model definition, a low posterior variance of the target and high sensitivity to the parameters values is required. Overparametrisations are then easily detected. Sensitivity analysis thus brings some insights into the analysis of model parsimony. Next, it enables to quantify of the importance of the model selection for characterising a target of interest. Some possible model-dependent features of the target can be highlighted; as a preferred feature, a low discrepancy between the output of different models seems reasonable. Overall, sensitivity analysis can be seen as a further tool for model assessment.

For our case study, the results showed that only a relatively low proportion of the uncertainty around unobserved components depends on the model used: the uncertainty about parameters had more important consequences than uncertainty around models. The effect of model uncertainty was mostly relevant regarding seasonal and the short-term movements of the series, while the trend was the component described in the most stable way over the three models. Next, the sensitivity analysis techniques pointed to model 2 as the model leading to a better definition of the series structure. This result did not agree with the BIC which favoured model 1 instead. It shows that sensitivity analysis complements other classical tools in a interesting way, shedding light about the accuracy of model definition with respect to the target of interest.

5 FURTHER INVESTIGATION OF THE ROLE OF SENSITIVITY ANALYSIS IN BUSINESS STATISTICS.

One of the typical problems tackled in econometrics is forecasting the future value of an endogenous variable given the past values of a number of exogenous ones. In the quite rare case of deterministic exogenous variables, the forecast is said to be unconditional, i.e. it is not conditioned by the probability density function (PDF) of the exogenous variables, given their deterministic character.
Unfortunately, in most econometric applications exogenous variables must be forecasted too, so the forecast becomes conditioned by the prediction of the exogenous variables. In Bayesian terms, any forecast based on a model that involves stochastic variables and/or parameters (from now on input factors) is conditioned by the input factors, the input factors being assigned a prior distribution. The stochastic nature of input factors has two effects on the reliability of forecasts. First, the variance of the estimate is higher since the variance of the inputs increases the variance of the conditional expectations (see Pindick and Rubinfeld, 1997 pp. 195-198 for a single regression example). Second, the choice of the most reliable forecasting model is far more difficult in the conditional case. A model that shows a good fit in terms of unconditional forecasting may show a poor performance in the conditional case, because of the variance introduced by the input factors. Often it may be complex to discriminate between several models with traditional econometric tests, because the PDF of the forecast in most econometric applications is not normally distributed. In particular it might be difficult to identify the relation between the variance of conditional expectation and the variance of any single input factor.
The Aim of the applications presented in this report is to show how variance-based sensitivity analysis could improve econometric model screening and the analysis of forecast reliability. Sobol' Method (see Section 2.2) is used to analyze the sensitivity of the conditional forecast of two classical case studies: Solow's classic (1957) study of the technical change in U.S. Economy and the consumption function of Klein's (1950) Model 1. The structure of application is then as follows. In the next section, the single regression application to Solow's classic is presented. In the following one, a model screening exercise on Klein's consumption function is presented. The last section gives the summary. The data used for the two application are taken from Greene (1997) 'Econometric Analysis'.

5.1 Sensitivity analysis of single regression model, Solow classic

Solow classic model suggests an aggregate production function where the aggregate output depends on capital and labor. One of the goals of his study was to define if aggregate output could be expressed as a function of labor and capital only, technological change being considered exogenous as stated in Neoclassic growth theory. Solow tested five different models; in this exercise we will introduce a sixth one. We will first screen among the six models with regression analysis on the sample 1909-1939; we decided to truncate the series in 1939 to avoid the II World War effect and the use of the dummy variable after 1943 as suggested by Greene (1997). We will then choose the one with the best fit according to econometric tests. Finally, we will forecast the outcome of the selected model for 1940 and perform a sensitivity analysis to define which input factors influence the most the variance of the expected output. Forecast and sensitivity analysis will be performed in a Monte Carlo framework, i.e. we will scan the entire PDF of model output instead of studying only the neighborhood of the central tendency.

5.1.1 Model selection

Solow suggested an aggregate production function defined implicitly as q(t)=A(t)f[k(t)] where q(t) is aggregate output per work-hour ($/hours), k(t) is the aggregate capital labor ratio, and A(t) is the technology index. We will estimate the five following explicit definitions, proposed by Solow, involving two coefficients b0 b1:

$$q/A = b_0 + b_1 * ln(k) \quad (27)$$

$$q/A = b_0 - b_1/k \quad (28)$$

$$ln(q/A) = b_0 + b_1 * ln(k) \quad (29)$$

$$ln(q/A) = b_0 - b_1/k \quad (30)$$

$$q(t)/A(t) - q(t-1)/A(t-1) = b_0 + b_1 * k(t-1) \quad (31)$$

We also tested a further model:

$$ln(q(t)/A(t)) - ln(q(t-1)/A(t-1)) = b_0 + b_1 * (ln(k(t)) - ln(k(t-1)) \quad (32)$$

Note that all models express a linear relationship between q/A or a transformation of it (natural logarithmic transformation, or first difference logarithmic transformation) and k or a transformation of it (natural logarithmic, first difference logarithmic or an inverse transformation).
All models can be estimated with a single regression. Results are given in Table 5.1:
It is easy to find in the set of six models the functional form that will give, in principle, the best forecasting results from the analysis of regression statistics. First of all Model (31) can be eliminated because of the poor goodness of fit statistics (R2, F-stat, etc.). All the other models seem acceptable in term of coefficients tests and goodness of fit, but once serial correlation is considered (Durbin Watson statistics), it becomes clear that models (28), (29) and (30) are strongly affected by 2nd order positive serial correlation and models (27) and (31) are affected too, albeit in weaker form. The best functional form is than given by model (32) even if model (27) shows slightly better results in term of Akaike and Schwarz information criterions.
The forecast of the q/A ratio will be then performed using model (32) where:

$$b_0 \sim N(\hat{b}_0 ' \hat{\sigma}_{b_0}) \quad (33)$$

and

$$b_1 \sim N(\hat{b}_1, \hat{\sigma}_{b_1}) \quad (34)$$

The last information required to forecast the ratio q/A in T+1 is the value of k in the same time period. We tried several model specifications, the best results were given by an auto-regressive model computed on the first difference of ln(k) (Table 5.2):

The value of k in T+1 can be computed using the model:

$$dln(k(t)) = a_1 dln(k(t-1)) \quad (35)$$

where dln(k(t)) is defined as:

$$dln(k(t)) = ln(k(t)) - ln(k(t-1)) \quad (36)$$

and $a_1 \sim N(\hat{a}_1, \hat{\sigma}_{a_1})$

We are now ready to forecast q/A in T+1=1940.

5.1.2 Montecarlo forecast and variance decomposition

The final form of the 1-step ahead forecasting model can be found by the substitution of (35) into (32) and some simplifications:

$$dln(q(t+1)/A(t+1)) = f(b_0, b_1, a_0) = b_0 + b_1 * (a_1 * dln(k(t))) \quad (37)$$

The traditional way of forecasting the value of dln(q(t)/A(t)) in T+1, would be to take the expected value of the left hand side of (37). Far more complex would be the task of drawing it's confidence intervals, because dln(q(t+1)/A(t+1)) is not normally distributed because of the product b1*a1. Pindick and Rubinfeld (1997) suggest an estimate of the confidence interval of the forecast:

1. Calculate the 95 percent confidence intervals associated with the forecast that could be obtained were we to select $\hat{X}_{t+1}$ to be two standard deviations higher or lower, i.e., the confidence intervals associated with $Y^{*}_{T+1} = \hat{\alpha} + \hat{\beta}(\hat{X}_{T+1} + 2\sigma_u)$ and $Y^{**}_{T+1} = \hat{\alpha} + \hat{\beta}(\hat{X}_{T+1} - 2\sigma_u)$.

2. The final interval prediction is taken to be the union of the two confidence intervals; i.e., it contains all the values of $\hat{Y}_{t+1}$ "…belonging to at least one confidence interval.

Another way to draw confidence intervals and analyze the conditional variance of forecast is to scan the complete distribution of the forecast via straightforward Montecarlo, generate a significant sample and then study how it is conditioned by the input factors. In order to apply Montecarlo and variance decomposition techniques, it must be observed that dln(q(t)/A(t)) is conditioned by the distribution of the m=3 input factors b0,b1 and a1.
Sobol' method requires two sample matrices. Hence to compute the conditional forecasts of y(t)=dln(q(t)/A(t)), we have sampled 2 matrices of column dimension N=10.000 from each of the three marginal distribution: [b0,b1,a1] and [b0',b1',a1']. The distribution of y(T+1)) corresponding to the first matrix is shown in figure 5.1.
The expected value of the distribution can be then computed as:

$$E[y(T+1)] \cong \hat{f}_0 = \frac{1}{N}\sum_{n=1}^{N} f(b_{0i}, b_{1i}, a_{1i}) = -0.0055 \quad (38)$$

and its total variance is:

$$Var[y(T+1))] \cong \hat{D} = \frac{1}{N}\sum_{n=1}^{N} f^2(b_{0i}, b_{1i}, a_{1i}) - \hat{f}_0^2 = 7.87E-07 \quad (39)$$

It is also possible to draw the confidence intervals without the normality constraint of traditional techniques, since it is possible to build a percentile distribution for any PDF. So via Montecarlo integrals it is possible both to forecast the value of y(T+1) and to build the confidence intervals.
The further step that can be achieved with Variance Based Methods is to define how the PDF of y(T+1) is conditioned by the distribution of the three input factors. Sobol' method in this specific case consists in the computation of 3 first order indices and 3 second order ones.
First conditional variances of first order $\hat{D}_i$, i=1,2,3, where the subscripts correspond respectively to b0, b1 and a1 must be computed:

$$\hat{D}_1 = \frac{1}{N}\sum_{n=1}^{N} f(b_0, b_1, a_1) f(b_0, b_1', a_1') - \hat{f}_0^2 \tag{40}$$

$$\hat{D}_2 = \frac{1}{N}\sum_{n=1}^{N} f(b_0, b_1, a_1) f(b_0', b_1, a_1') - \hat{f}_0^2 \tag{41}$$

$$\hat{D}_3 = \frac{1}{N}\sum_{n=1}^{N} f(b_0, b_1, a_1) f(b_0', b_1', a_1) - \hat{f}_0^2 \tag{42}$$

First order conditional variances $\hat{D}_i$ are computed as the difference between the squared expected value $\hat{f}_0^2$ and the sum of products of two vectors of forecasted values. The first vector is computed using the first sampled matrix [b0,b1,a1], the second one modifying the second matrix [b0',b1',a1']. The modification consists in the substitution of one of the sampled column vectors of the second matrix with the corresponding one of the first. So the variance of the forecast conditioned by the first input factor b0 is computed using the first sample matrix and the modified matrix [b0,b1',a1'], the variance of the second factor b1 is computed using the first matrix and the modified [b0',b1,a1'] and the variance of the third factor a1 is computed using the first matrix and the modified [b0',b1',a1]. The rational of the procedure is that computing the variance of the forecast conditioned by an input factor, consists in computing the variance of the forecast given the value of the conditioning input factor, hence keeping its value constant.
Once first order conditional variances have been computed, second order conditional variances are computed in a similar way:

$$\hat{D}_{12} = \frac{1}{N}\sum_{j=1}^{n} f(b_0, b_1, a_1) f(b_0, b_1, a_1') - \hat{f}_0^2 - \hat{D}_1 - \hat{D}_2 \tag{43}$$

$$\hat{D}_{13} = \frac{1}{N}\sum_{j=1}^{n} f(b_0, b_1, a_1) f(b_0, b_1', a_1) - \hat{f}_0^2 - \hat{D}_1 - \hat{D}_3 \tag{44}$$

$$\hat{D}_{23} = \frac{1}{N}\sum_{j=1}^{n} f(b_0, b_1, a_1) f(b_0', b_1, a_1) - \hat{f}_0^2 - \hat{D}_2 - \hat{D}_3 \tag{45}$$

Second order conditional variances $\hat{D}_{ij}$ measure the interaction effect of factor i with factor j: the variation of f(b0,b1,a1) due to factor i and j which cannot be explained by the sum of the individual effects of the two factors. They are computed keeping factor i and factor j constant (same vector column for factor i and j in both matrices) and subtracting the first order effects $\hat{D}_i$ and $\hat{D}_j$

Dividing by the total variance $\hat{D}$ as computed in (39), we get Sobol' indices of first and second order:

$$\hat{S}_1 = \frac{\hat{D}_1}{\hat{D}} = 0.0046 \tag{46}$$

$$\hat{S}_2 = \frac{\hat{D}_2}{\hat{D}} = 0.0007 \tag{47}$$

$$\hat{S}_3 = \frac{\hat{D}_3}{\hat{D}} = 0.9833 \tag{48}$$

$$\hat{S}_{12} = \frac{\hat{D}_{12}}{\hat{D}} = 0.0000 \tag{49}$$

$$\hat{S}_{13} = \frac{\hat{D}_{13}}{\hat{D}} = 0.0000 \tag{50}$$

$$\hat{S}_{23} = \frac{\hat{D}_{23}}{\hat{D}} = 0.0000 \tag{51}$$

Results are summarized and charted in Figure 5.2.
The most evident result is that almost all the variance (98%) of the forecast depends on the variance of factor 3: the coefficient, a1 of the autoregressive process of dln(k(T+1)). The other indices, first and second order, does not exceed 1%.
The overall result of the analysis is that the variance of the forecast depends almost exclusively on the volatility of the prediction of the capital labor ratio, while the production function plays almost no role in the explanation of the variance of the conditional forecast. So Solow' production function performed extremely well in the unconditional case, while Sobol' analysis pointed out that in the conditional case the performance is rather poor. In economic terms, we are in the situation in which the model itself is perfect, but it would be better not to base your decisions on it. Nor are we in the condition to choose a better forecasting model since all the proposed models present only the capital labor ratio as exogenous variable. The choice of another model will simply increase the variance of the output. So always from an economic point of view we can state that the growth of production cannot expressed as a function of the capital labor ratio, because we would not be able to predict the evolution of production, so Neoclassic growth theory does not hold. It must be mentioned that we could find the same results with the analysis of regression statistics. Consider the sum of square residuals of model (32) and of the forecasting process of dln(k(T+1)), 2.81E-05 and 0.037817 respectively. It is clear that the most of the variance of a model combining the two will depend mostly on the variance of the second.
So which is the value added of sensitivity analysis?
First of all it quantifies the share of conditional variance that can be accounted to any input factor. Second Montecarlo based tests are not constrained by the normality of the PDF of the outputs. Finally they are of great help in model screening, as it will become clear in the test case exposed in next paragraph.

5.2 Sensitivity analysis of multiple regression model, linear versus non linear modelling; Klein1 consumption equation

Klein (1950) Model 1 is a classic example of simultaneous equation model. Here we will focus only on the consumption function:

$$C_t = \alpha_0 + \alpha_1 P_t + \alpha_2 P_{t-1} + \alpha_3 (W_t^p + W_t^g) \tag{52}$$

Equation (52) express Consumption C_t, as a linear function of prices P_t and P_{t-1}, private wages W_t^p and government wages W_t^g. Klein estimated the parameters using U.S. data for 1921 to 1941. We decided, once more, to truncate the series (the data corresponding to 1941 clearly show break comparing to the previous trend). Again we will try to identify the forecasting model that gives the more reliable results, but we will treat a different situation.
First we will consider a situation that involves uncertainty not only in the model functional form, but also in the selection of variables that should be included in the model. Second we will consider a forecasting horizon of 5 years and we will compute Sobol' indexes for each acceptable model in each time period.

5.2.1 Model selection

In paragraph 2 we analyzed several models, but all of them were linear models of a transformation of the capital/labor ratio. Here we considered all the combination of the three variables proposed by Klein and several transformations of them (first difference, first logarithmic difference), we also estimated non-linear models such as logarithmic, exponential and others non linear models. While for Solow production function model (32) gave far better results then the other models, for Klein 1 we found three acceptable models:

$$C_t = b_{01} + b_{11}P_{t-1} + b_{21}W_t^P \tag{53}$$

$$dC_t = b_{02} + b_{12}dP_t + b_{22}d(W_t^p + W_t^g) \tag{54}$$

$$C_t = b_{03} + P_{t-1}^{b_{13}} + (W_t^p + W_t^g)^{b_{23}} \tag{55}$$

Estimation Statistics are given in Table 5.3.
All three models are acceptable in terms of goodness of fit, serial correlation and Schwarz information criterion. The three models are rather different. The first one does not include government wages, the second one is defined on the first difference of all variables and the third is an inherently non-linear model where the parameters appear as exponents. As for Solow model all parameters follow a normal distribution (model (55) has been estimated via an iterative linearization method, so b13 and b23 too follow a normal distribution (See Pindick and Rubinfeld (1997) pp. 235-238).
Concerning the forecast of the exogenous variables the three following models gave the better results according to Scharwz criterion:

$$dlln(P_t) = a_{01} + a_{11}dln(P_{t-1}) \tag{56}$$

$$dlln(W_t^p) = a_{02} + a_{12}dln(W_{t-1}^p) \tag{57}$$

$$d(W_t^g) = a_{03} + a_{13}d(W_{t-1}^g) \tag{58}$$

Estimation outputs are reported in Table 5.4:
We are ready to perform the conditional forecast.

5.2.2 Montecarlo forecast and variance decomposition

The three forecasting models are obtained by the substitution of equations (56), (57) and (58) in equation (53), (54) and (55), leading to:

$$C_t = f(b_{01}, b_{11}, b_{21}, a_{01}, a_{11}, a_{02}, a_{12}) = b_{01} + b_{11}\left(e^{a_{01}}\frac{P_{t-2}^{a_{11}+1}}{P_{t-3}^{a_{11}}}\right) + b_{21}\left(e^{a_{02}}\frac{W_{t-1}^{P\,a_{12}+1}}{W_{t-2}^{P\,a_{12}}}\right) \tag{59}$$

$$dC_t = f(b_{02}, b_{12}, b_{22}, a_{01}, a_{11}, a_{02}, a_{12}, a_{03}, a_{13}) =$$
$$= b_{02} + b_{12}\left(e^{a_{01}}\left(\frac{P_{t-1}^{a_{11}+1}}{P_{t-2}^{a_{11}}} - \frac{P_{t-2}^{a_{11}+1}}{P_{t-3}^{a_{11}}}\right)\right) + b_{22}\left(e^{a_{02}}\left(\frac{W_{t-1}^{P\,a_{12}+1}}{W_{t-2}^{P\,a_{12}}} - \frac{W_{t-2}^{P\,a_{12}+1}}{W_{t-3}^{P\,a_{12}}}\right) + a_{03} + a_{13}d(W_{t-1}^g)\right) \tag{60}$$

$$C_t = f(b_{03}, b_{13}, b_{23}, a_{01}, a_{11}, a_{02}, a_{12}, a_{03}, a_{13}) =$$

$$= b_{03} + \left(e^{a_{01}} \frac{P_{t-2}^{a_{11}+1}}{P_{t-3}^{a_{11}}} \right)^{b_{13}} + \left(\left(e^{a_{02}} \frac{W_{t-1}^{P\,a_{12}+1}}{W_{t-2}^{P\,a_{12}}} \right) + \left(a_{03} + (a_{13}+1) W_{t-1}^{g} - a_{13} W_{t-2}^{g} \right) \right)^{b_{23}} \quad (61)$$

where all a.. and b.. are normally distributed with the means and variances given in table 5.3 and 5.4. Note that the values of prices and wages at the previous times periods are allways obtained from equations (56), (57) and (58), so the input factors are the parameters a.. and b.. while prices and wages enter only as a deterministic constant valued at the last time period: 1939-1940 for both wages variable and 1938-1939 for prices values. It also rather evident that any forecast obtained by the three model does not present a normal distribution since they all involve products, exponential and even powers of normally distributed factors. So the goodness of fit statistics of the model estimates clearly does not apply in the conditional forecast case.

The output of models (60) and (61) are conditioned by the PDF of 9 input factors, while model (59) forecast depends only on 7 factors since it does not include government wages. If we computed Sobol indexes for all the 7 or 9 orders, we should have computed 2k-1=127 indexes for model (59) and 2k-1=517 for models (60) and (61), far two many for our purpose. We decided to compute only the indexes of the first order, so 7 for model (59) and 9 for the other two models. Alternatively one could compute the total sensitivity indexes (see Section 2.2).

Results are reported in Figure 5.3.

Where:

S1 , S2 and S5 correspond to b0., b1. and b3. ;

S3 and S4 to a01 and a11;

S6 and S7 to a02 and a12;

S8 and S9 to a03 and a13.

Several interesting observation can be made.

First of all government wages' variable could be taken as a constant since it never plays a relevant role in conditioning the variance of the forecast, S8 and S9 never exceed 10-4. That indicates that the volatility of the factors associated with the variable is much lower than the one of the others.

Model (59) and (61) show quite similar results. In the first period a high proportion of the variance of the forecast depends on b3., the parameter associated to the wages variable S5=0.45 for model (61) and 0.51 for (59). Then their importance progressively decreases, while the importance of S6 and S7 increases, corresponding to the factors that give the forecast of private wages. In 1945 the most of the variance depends on them. The phenomenon has an economical explanation; that indicates that the elasticity of consumption to wages does not involve a great volatility through time (see b2. coefficient statistics), but the variance of the forecast of private wages increases quite fast moving from the last time period. Finally observe that the variance of the factors associated to prices plays a very little role (S3 and S4 never exceed 0.07 and decreases moving from 1940 on).

Model (60) has quite a different behavior. The main difference can be easily observed with the graphs. While in Model (59) and (61) the importance of S5 decreases through time while the importance of S6 and S7 increases; in model (60) the relative importance of factors tends to be rather constant. Farther more the importance of input factors is more spread among factors, no factor conditions more than 0.30 of the overall variance in any time period. Finally while in the other models price factors did not play any role in model (60), S3 conditions between 0.21 and 0.17. In economics terms that means that the variance of first difference of consumption is sensitive to the first difference of prices, while the variance of consumption is not sensitive to the variance of prices; i.e. consumption is not conditioned by the level of prices, but by the rate of inflation. Another important difference is that higher order indices in model (60) have an increasing role through time (0.23 in the last forecasting period) that indicates that some joint conditional variance effects are present, so total sensitivity indexes are more suitable to analyze model (60).

All that considered it is clear that the more suitable model to forecast consumption is model (60). The main reason is that the variance of forecast is more spread among different factors. The variance of forecast in model (59) and (61) depends only on factors related to private wages, so an unexpected change in private wages would involve a rather serious error of forecast. The variance in model (59) depends on several factors related both to private wages and prices, so the chance of getting an unreliable forecast is less likely; the two variable are orthogonal, an unexpected change in both might append only by chance. In some way Sobol' analysis can be compared to a financial risk analysis: spreading the risk of forecast error among

different variables reduces the risk of unreliable forecasting results. Note the no such information is given by standard econometric techniques: model (59), (60) and (61) show very little differences in term of regression statistics.

5.3 Conclusions for the econometric test case

In the present application we presented a first attempt to apply Variance Based Sensitivity Methods to conditional forecasting, the results can be summarized as follow.
Compared to traditional techniques, Montecarlo forecasting combined with sensitivity analysis shows three improvements.
First using Sensitivity analysis it is straight foreword to find out if a model that performed well in an unconditional case, gives valuable results also in the conditional case (see paragraph 2.2).
Second Monte Carlo based forecasts are not constrained by the non-normality of the PDF of outputs. Confidence intervals can be designed from the percentile distribution of the forecast.
Third sensitivity analysis can be extremely useful in model screening exercises (see paragraph 3.2) when several models seem to perform at the same level. Sensitivity analysis gives a better insight to conditional variance, hence it eases the choice of the model that presents the more steady conditioning of variance.

6 CONCLUSIONS

The present report is based on testing SA methodologies on selected cases relevant to official statistics. The first example related to the use of environmental indices for policy-making. Uncertainty analysis has been successfully employed for testing the robustness of the process when various sources of uncertainty (the factors) play simultaneously. An innovative method of sensitivity analysis, developed at the JRC, has been used to reveal at what extent the factors affect the pressure-to-decision index. The sensitivity analysis is used for grouping the factors according to different logical structures. This tool appears adequate in helping the analyst reducing the uncertainty in the model output below reasonable levels. This increases the likelihood that a certain policy could be implemented.

The second example, on time series analysis, has shown that SA methods are of great help in model screening, as they provide an assessment of how “balanced” a model calibration is given the selected model, the available data and the set of parameters to be calibrated.

The final test case showed that using Sensitivity analysis one can verify whether a model that performed well in an unconditional case, gives valuable results also in the conditional case.

The report also discusses the similarities and differences between the statistical concept of **robustness** and sensitivity analysis, by applying SA to a classical example from Huber (see section 2.3).

A general conclusion of the present work is that, especially when the model is used to drive a choice or a decision, the importance of the associated uncertainties should be quantified, and the relevance of the model ensured. Quantitative global sensitivity analysis could be of use in this context.

Very often the use of a model is forced within an optimisation context, where the issue is to find some least square solution to estimate the input factors which disagree least with observations. Any model must include false and/or unrealistic assumptions in order to be of practical use, and these will not be made explicit in the calibration process. We see with empathy any attempt of mapping the input factors space into the prediction space, including alternative model structures, or the bootstrapping of the modelling process, and any reverse process, e.g. the mapping of the plausible outcomes back into the inputs space. SA could be an element of this alternative approach to “calibration”.
There is considerable scope for improvement both in the extension and depth of application of SA. An important technicality is that in order to use quantitative techniques (even to use screening methods more complicate than elementary OAT), the model must be interfaced with some kind of “driver” programme to import the sampled data into the model. In turn, the model output must then be exported to the algorithm used to compute sensitivities.

Examples of instrumental use of models can be found in the literature, especially when models are used for making decisions having a large social and economic impact. Thus, it is not surprising to meet cynic opinions about models. An example is in The Economist, (1998) where one reads that
"based largely on an economic model […] completing K2R4 [a nuclear reactor] in 2002 has a 50% chance of being 'least cost' "
Given that the model was used, to contradict a panel of experts on the opportunity to build the aforementioned reactor, The Economist comments:
"Cynics say that models can be made to conclude anything provided that suitable assumptions are fed into them."
In the spirit of the approach defended in this report, it would be highly instructive to look at what factors were determining the variation around the 'least cost' region. The outcome of this analysis could then provide experts with additional insight.
A possible cause of frustration in SA is the large volume of data generated by the analysis in the presence of large multivariate input and output. We believe that the analysis can be made more effective by focusing not on the model output but on the statement that such output is supposed to serve. To this effect, "model use" should be declared before the UA and SA is performed.

The global quantitative SA methods discussed here can adequately characterise the model behaviour, and thus justify a more ambitious role of SA, especially in the epistemic perspective. Hence, model complexity and model corroboration should be presented as conflicting, e.g. the more complex a model, the less likely its corroboration. Increasing computer capacity on one hand, and epistemic awareness on the other, suggest that model's relevance become a prescription for the correct use of models.

7 REFERENCES

Adriaanse A., 1993, Environmental Policy Performance Indicators, *Sdu Uitgeverij Koninginnegracht,* ISBN 90 12 08099 1.

Archer, G.E.B., A. Saltelli, and I.M. Sobol, 1997, Sensitivity Measures, ANOVA-like Techniques and the use of Bootstrap', *Journal of Statistical Computations and Simulations*, **58**, 99-120.

Beck M. B., J. R. Ravetz, L. A. Mulkey, T. O.Barnwell, 1997, On the problem of model validation for predictive exposure assessments, *Stochastic Hydrology and Hydraulics*, **11**, 229-254.

Bell, W.R., 1989, Comment to ``Sensitivity Analysis of Seasonal Adjustments: Empirical Case Studies", by J.B. Carlin and A.P. Dempster, *Journal of American Statistical Association,* **84** (405) 22-24.

Bell W.R., and S.C. Hillmer, 1984, Issues Involved with the Seasonal Adjustment of Economic Time Series, *Journal of Business and Economic Statistics*, **2**, 291-320.

Bell W.R., and M.C. Otto, 1993, Bayesian Assessment of Uncertainty in Seasonal Adjustment with Sampling Error Present, SRD Research Report number RR-92/12, Washington, D.C: U.S. Bureau of the Census.

Bowman, K.O. and L.R. Shenton, 1975, Omnibus Tests Contours for Departures from Normality Based on $\sqrt{b_1}$ and b_2, *Biometrika*, **62**, 243-250.

Box, G.E.P and G.M Jenkins, 1970, *Time Series Analysis: Forecasting and Control*, San Francisco: Holden Day.

Box, G.E.P and Tiao, G.C, 1973, *Bayesian Inference in Statistical Analysis*,.Reading, Massachussetts: Addison-Wesley.

Burman, J.P., 1980, Seasonal Adjustment by Signal Extraction, *Journal of the Royal Statistical Society*, Ser. A, **143**, 321-337.

Brockwell, P.J, and R.A. Davis, 1991, *Time Series: Theory and Methods*, 2nd ed., Springer Verlag: New York.

Carlin, J.B. and A.P. Dempster, 1989, Sensitivity Analysis of Seasonal Adjustments: Empirical Case Studies, *Journal of the American Statistical Association*, **84** (405), 6-20.

Chan, K., S. Tarantola, and F. Campolongo, Eds., 1998, Proceedings of the Second International Symposium on Sensitivity Analysis of Model Output (SAMO98), Venice (I), EUR report 17758 EN, Luxembourg.

Computer Physics Communications, 1999, Special Issue on sensitivity analysis, Saltelli, Chan and Scott Eds., **117**(1-2).

Chatfield C., 1993, Model uncertainty, data mining and statistical inference, *Journal of Royal Statistical Society A*, **158** (3) , 419-466.

Cukier, R. I., H. B. Levine, and K. E. Schuler, 1978, Nonlinear sensitivity analysis of multiparameter model systems, *Journal of Computational Physics*, **26**, 1-42.

Daniel, C., 1958, On varying one factor at a time, *Biometrics*, **14**, 430-431.

Daniel, C., 1973, One-at-a-Time-Plans, *Journal of the American Statistical Association*, **68**, 353-360.

Draper, D., 1995, Assessment and Propagation of Model Uncertainty, *Journal of the Royal Statistical Society*, Ser. B, **57** (1), 45-97.

Draper, D., A. Pereira, P. Prado, A. Saltelli, R. Cheal, S. Eguilior, B. Mendes, and S. Tarantola, 1999, Scenario and parametric uncertainty in GESAMAC: a methodologies study in nuclear waste disposal risk assessment, *Computer Physics Communications*, forthcoming.

European Environment Agency, 1996, Atmospheric Emission Inventory Guidebook by EMEP/CORINAIR 1996. Published by the European Environment Agency, Copenhagen.

Eurostat, 1999, Towards Environmental Pressure Indicators for the EU. Published by Eurostat.

Fedra, K., G. Van Straten, M. B. Beck, 1981, Uncertainty and Arbitrariness in Ecosystems Modeling: a Lake Modeling Example", *Ecological Modelling*, **13**, 87-110.

Fisher, R.A., 1935, *Design of Experiment*. Olivier and Boyd, Edinburgh.

Fuller, W.A., 1996, *Introduction to Statistical Time Series*, 2nd ed., Wiley: New York.

Geweke, J., 1996, Monte Carlo Simulation and Numerical Integration, in *Handbook of Computational Economics*, vol. 1, ed. H.M. Amman, D.A. Kendrick, and J. Rust, Elsevier Science.

Gilks, A., Richardson, S. and Spiegelhalter D.J., 1996, Introducing Markov Chain Monte Carlo, in *Monte Carlo Markov Chain in Practice*, ed. by Gilks, A., Richardson, S. and Spiegelhalter D.J. London: Chapman & Hall.

Goldsmith, C. H., 1998, Sensitivity Analysis, in *Encyclopedia of Biostatistics*, by Peter Armitage (Editor), Theodore Colton (Editor), John Wiley & Sons.

Gómez, V. and A. Maravall, 1996, Programs SEATS and TRAMO: Instructions for the User, Working Paper No. 9628, Bank of Spain.

Greene W. H. 1997, Econometric Analysis, Prentice-Hall.

Hannan, E.J., 1980, The Estimation of the Order of an ARMA Process, *Annals of Statistics*, **8** (5), 1071-1081.

Harvey, A.C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press.

Helton, J. C., 1993, Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal, *Reliability Engineering and System Safety*, **42**, 327-367.

Helton, J. C. and D. E. Burmaster, 1996, Treatment of aleatory and epistemic uncertainty in performance assessments for complex systems", Guest editorial, Special Issue of *Reliability Engineering and System Safety*, **54**, 91-94.

Hillmer S.C. , and G.C. Tiao, 1982, An ARIMA-Model-Based Approach to Seasonal Adjustment, *Journal of the American Statistical Association*, **77**, 63-70.

Homma T., and A. Saltelli, 1996 , Importance measures in global sensitivity analysis of model output, 1996, *Reliability Engineering and System Safety*, **52** (1) 1-17.

Houghton J.T., L.G. Meira Filho, B.A. Callander, N. Harris, A. Kattenberg and K. Maskell Eds., 1996, Climate Change 1995, The Science of Climate Change, Contribution of WGI to the Second Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge.

Huber P., 1981, *Robust Statistics.* Wiley: New York.

Iman, R. L. and J. C. Helton, 1988, An investigation of uncertainty and sensitivity analysis techniques for computer models, *Risk Analysis*, **8** (1), 71-90.

Janssen, P. H. M., W. Slob, and J. Rotmans, 1990, Uncertainty analysis and sensitivity analysis: an inventory of ideas, methods and techniques from the literature, *RIVM-report, 958805001,* Bilthoven, the Netherlands.

Jesinghaus J., 1999, A European System of Environmental Pressure Indices. First Volume of the *Environmental Pressure Indices Handbook*.

Journal of Multicriteria Decision Analysis, 1999, Special Issue on sensitivity analysis, D. Rios Insua ed., forthcoming.

Journal of Statistical Computation and Simulation, 1997, Special Issue on Sensitivity Analysis, **57** (1-4).

Kass, R.E, and A.E. Raftery, 1995, Bayes Factors, *Journal of the American Statistical Association*, **90** (430), 773-795.

Klein L., 1950, Economic Fluctuations in the United States 1921-1941. New York: John Wiley and Sons.

Maravall, A. and Planas, C., 1999, Estimation Errors and the Specification of Unobserved Component Models, forthcoming in *Journal of Econometrics*.

McKay, M. D., 1995, Evaluating Prediction Uncertainty, *NUREG/CR-6311, U.S. Nuclear Regulatory Commission and Los Alamos National Laboratory*.

NRW State Environment Agency, 1997, Identification of Relevant Industrial Sources of Dioxins and Furans in Europe. Publication of the LandesUmweltAmt - North-Rhine-Westphalia State Environment Agency, Essen (D)

Oreskes, N., K. Shrader-Frechette, K. Belitz, 1994, Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences, *Science*, **263**, 641-646.

Pastres, R., D. Franco, G. Pecenik, C. Solidoro, and C. Dejak, 1997, Local sensitivity analysis of a distributed parameters water quality model, *Reliability Engineering and System Safety*, **57**, 21-30.

Pierce, D.A. 1978, Seasonal Adjustment when Both Deterministic and Stochastic Seasonality are Present, in *Seasonal Analysis of Economic Time Series*, ed. A. Zellner, Washington, D.C.: U.S. Dept. of Commerce, Bureau of the Census, 242-269.

Pindick R. and Rubinfeld D. 1997, Econometric Models & Economic Forecast, McGraw-Hill.

Puolamaa M., M. Kaplas, T. Reinikainen, 1996, Index of Environmental Friendliness – A Methodological Study, *Statistics Finland*, ISBN 951-727-272-3.

Rabitz, H., 1989, System analysis at molecular scale, *Science*, **246**, 221-226.

Rabitz, H., O. F. Alis, , J. Shorter, and K. Shim, 1999, Efficient input-output model representations, *Computer Physics Communications*, special issue on sensitivity analysis, forthcoming.

Reliability Engineering and System Safety, 1997, Special Issue on Sensitivity Analysis, **57** (1).

Rose, K. A., E. P. Smith, R. H. Gardner, A. L. Brenkert, S. M. Bartell, 1991, Parameter Sensitivities, Monte Carlo Filtering, and Model Forecasting under Uncertainty, *Journal of Forecasting*, **10**, 117-133.

Rosen, R., 1991, A Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life, Columbia University Press, New York.

Saltelli, A., and T. Homma, 1992, Sensitivity analysis for model output; Performance of black box techniques on three international benchmark exercises", *Computational Statistics and Data Analysis*, **13** (1), 73-94.

Saltelli, A., T.H. Andres, and T. Homma, 1993, Sensitivity Analysis of Model Output: An Investigation of New Techniques, *Computational Statistics and Data Analysis*, **15**, 211-238.

Saltelli, A., F. Campolongo, and S. Tarantola, 1999, Sensitivity analysis as an ingredient of modelling, (Part II) . Submitted to *Statistical Science*.

Saltelli, A., S. Tarantola, and F. Campolongo, 1999, Variance based measures in sensitivity analysis (Part I). Submitted to *Statistical Science*.

Saltelli A., S. Tarantola, and K. Chan K., 1999, A quantitative, model independent method for global sensitivity analysis of model output, *Technometrics*, **41**,1, 39-56.

Sobol', I. M., Sensitivity estimates for non-linear mathematical models, *Matematicheskoe Modelirovanie*, **2**, 112-118, 1990, translated as: I. M. Sobol', 1993, Sensitivity analysis for non-linear mathematical models, *Mathematical Modelling & Computational Experiment*, **1**, 407-414.

Solow R. 1957, 'Technical Change and the Aggregate Production Function', Review of Economics and statistics, 39 pp. 312-320.

Spear, R.C., 1970, The Application of Kolmogorov-Rényi Statistics to Problems of Parameter Uncertainty in Systems Design, *International Journal of Control*, **11** (5), 771-778.

The Economist, 1998, “More fall-out from Chernobyl”, June 27th, p. 98.

Tomovic, R., and M. Vukobratovic, 1972, General Sensitivity Theory, New York, NY: Elsevier.

Tong, H., 1990, *Non Linear Series: A Dynamical Systems Approach*, Oxford University Press, Oxford.

Turanyi, T., 1990, Sensitivity analysis of complex kinetic systems. Tools and applications, *Journal of Mathematical Chemistry*, **5**, 203-248.

Young, P. C., S. D. Parkinson, M. Lees, 1996, Simplicity out of complexity: Occam’s razor revisited, *Journal of Applied Statistics*, **23**, 165-210.

Young, P., 1999, Data-based Mechanistic Modelling, Generalised Sensitivity and Dominant Mode Analysis, *Computer Physics Communications,* **117**, 113-129.

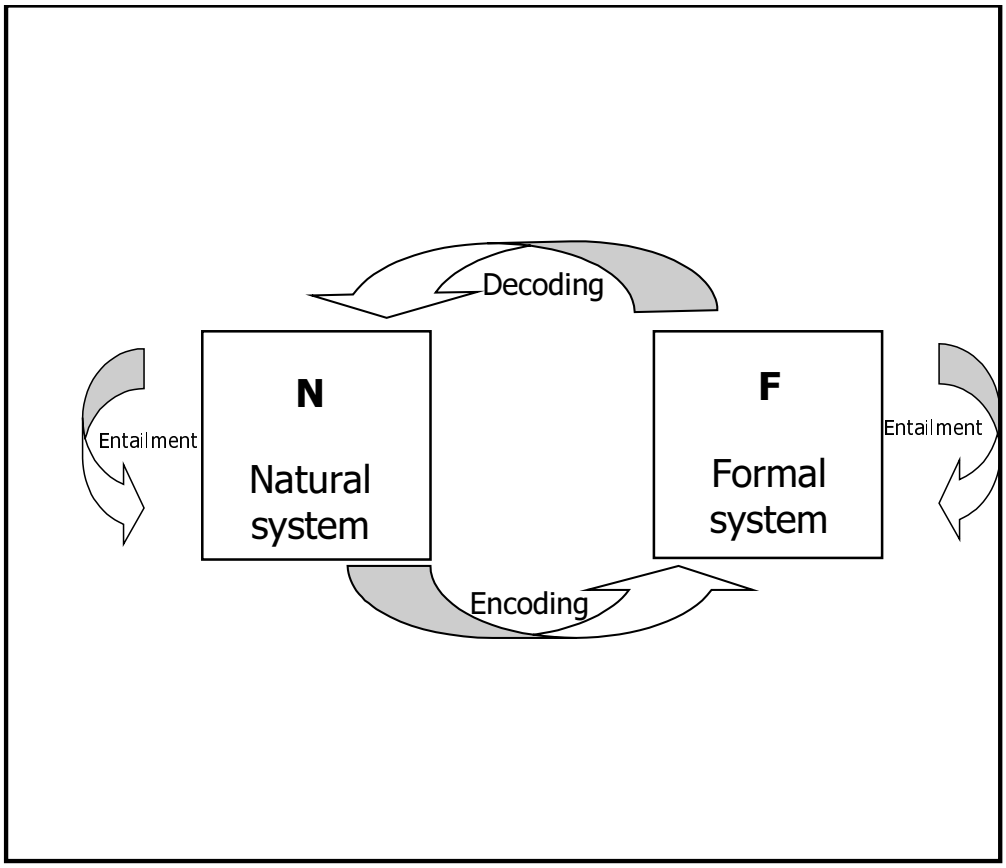


Figure 2.1: Rosen's formalisation of the scientific process

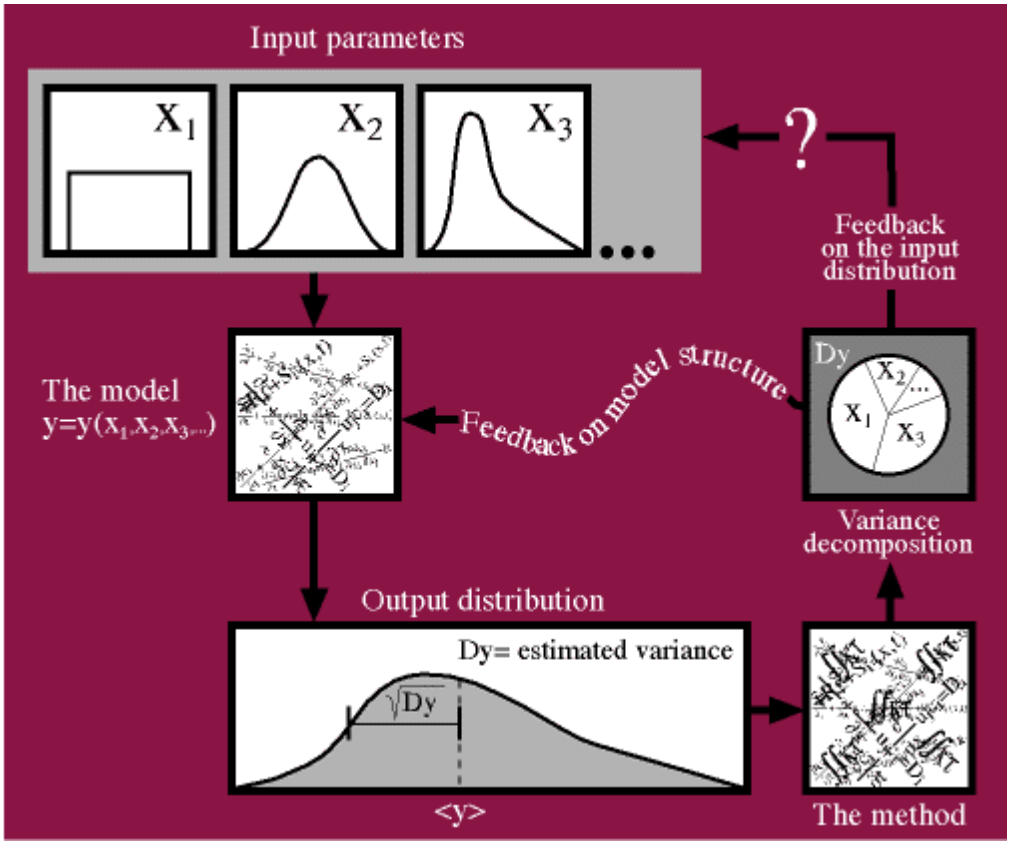


Figure 2.2: General scheme of Sensitivity Analysis

Air Pollution	Emissions of nitrogen oxides (NOx)	Emissions of NMVOC	Emissions of sulphur dioxide (SO2)	Emissions of particles	Consumption of gasoline & diesel oil by road veh.s	Primary energy consumption
Climate Change	Emissions of carbon dioxide (CO2)	Emissions of methane (CH4)	Emissions of nitrous oxide (N2O)	Emissions of chloro-fluoro-carbons	Emissions of nitrogen oxides (NOx)	Emissions of sulphur oxides (SOx)
Loss of Biodiversity	Protected area loss, damage and fragmentation	Wetland loss through drainage	Agriculture intensity: area used for intensive..	Fragmen-tation of forests & landscapes	Clearance of natural & semi-nat.l forests	Change in traditional land-use practice
Marine Environment & Coastal Zones	Eutrophi-cation	Overfishing	Develop-ment along shore	Discharges of heavy metals	Oil pollution at coast & at sea	Discharges of halogena-ted organic compounds
Ozone Layer Depletion	Emissions of chloro-fluoro-carbons	Emissions of bromo-fluoro-carbons	Emissions of hydro-chlorofluoro-carbons	Emissions of nitrogen oxides (NOx)	Emissions of chlorinated carbons	Emissions of methyl bromide (CH3Br)
Resource Depletion	Water consumption per capita	Use of energy per capita	Increase in territory permanently occupied	Nutrient balance of the soil	Electricity production from fossil fuels	Timber balance (new growth/ harvest)
Dispersion of Toxic Substances	Consumption of pesticides	Emissions of persistent organic pollutants	Consumption of toxic chemicals	Index of heavy metal emissions to water	Index of heavy metal emissions to air	Emissions of radioactive material
Urban Environmental Problems	Energy consumption	Non-recycled municipal waste	Non-treated wastewater	Share of private car transport	People endangered by noise emissions	Land use (change from natural to built-up..)
Waste	Waste landfilled	Waste incinerated	Hazardous waste	Municipal waste	Waste per product	Waste recycled/ material recovered
Water Pollution & Water Resources	Nutrient use (nitrogen & phosphorus)	Ground water abstraction	Pesticides used per hectare of agric.l area	Nitrogen used per hectare of agric.l area	Water treated/ water collected	Emissions of organic matter as BOD

Figure 3.1: Eurostat environmental pressure indicators, from (Eurostat (1999)). In the first column the ten environmental themes are given. For each environmental theme the six most important indicators, ranked from left to right, are reported along each single row.

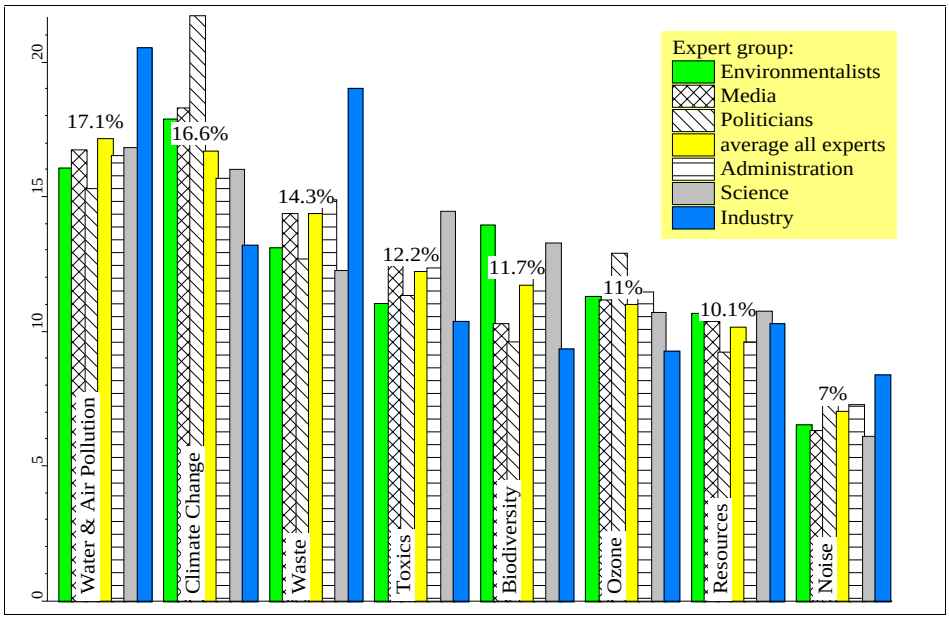


Figure 3.2: Overall importance of environmental themes (RSD 1990). In the German study, over 600 experts had been asked how they perceive the overall importance of various environmental themes. “Ordinary” water and air pollution had been sub-summed into one theme. To enable comparisons to the Finnish and Eurostat indicator sets, the result (17%) has been split artificially on the basis of the water:air relation used by Adriaanse for the Dutch pressure index (Water pollution was not considered in our study, due to lack of data).

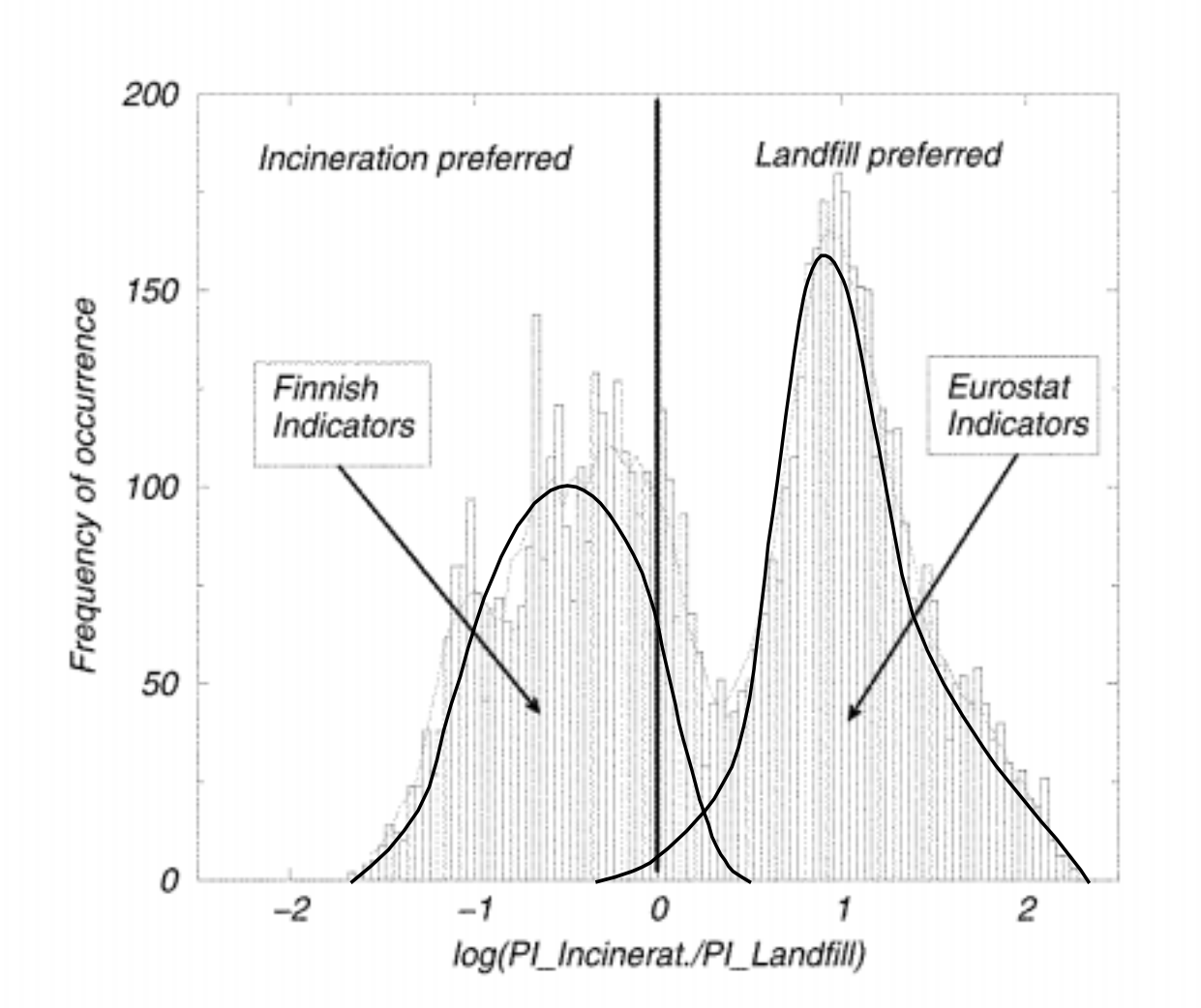


Figure 3.3: Results of UA for the exercise where the indicators of eco-toxicological effect (ET) and of Dispersion of Toxics (DT) are not included.

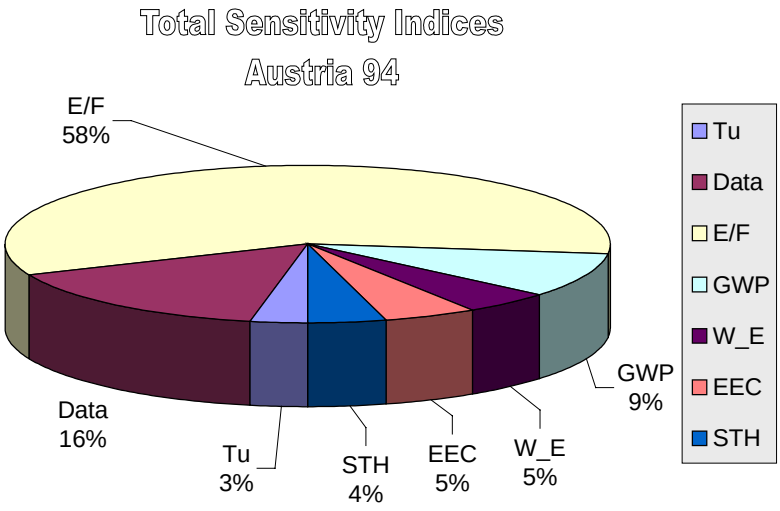


Figure 3.4. Results of SA for the first exercise. Both the Finnish and the Eurostat sets of indicators have been used (factor [E/F]). Uncertainty on indicators is much larger than that of the available data.

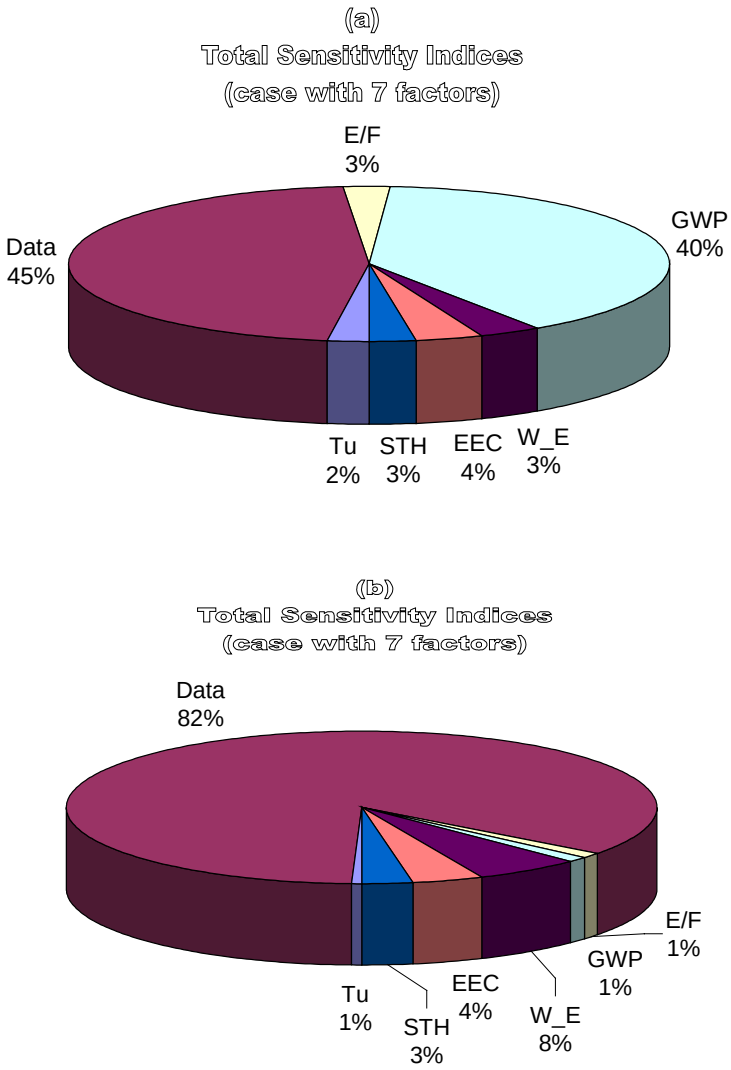


Figure 3.5: Results of SA for the second exercise using the Finnish indicators (a) and the Eurostat indicators (b).

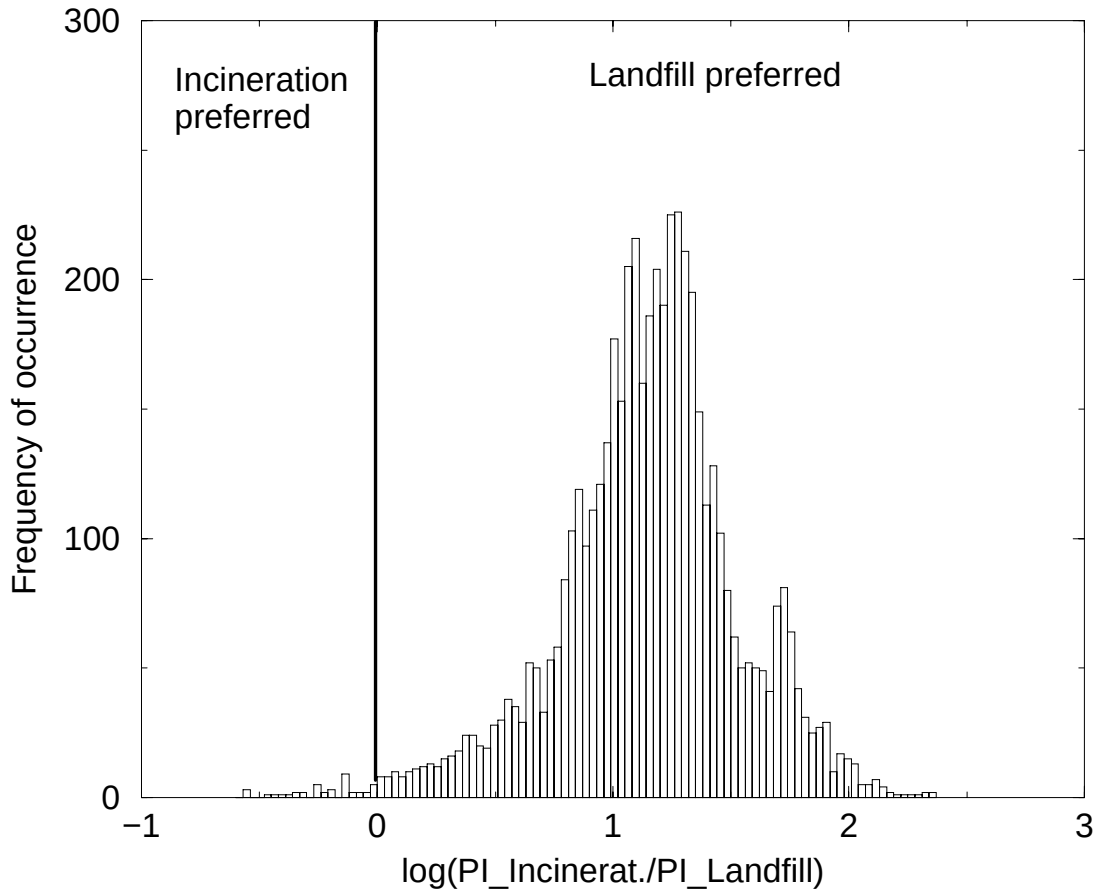


Figure 3.6: Results of UA for the third exercise, where the indicators of eco-toxicological effect (ET) and of Dispersion of Toxics (DT) are included.

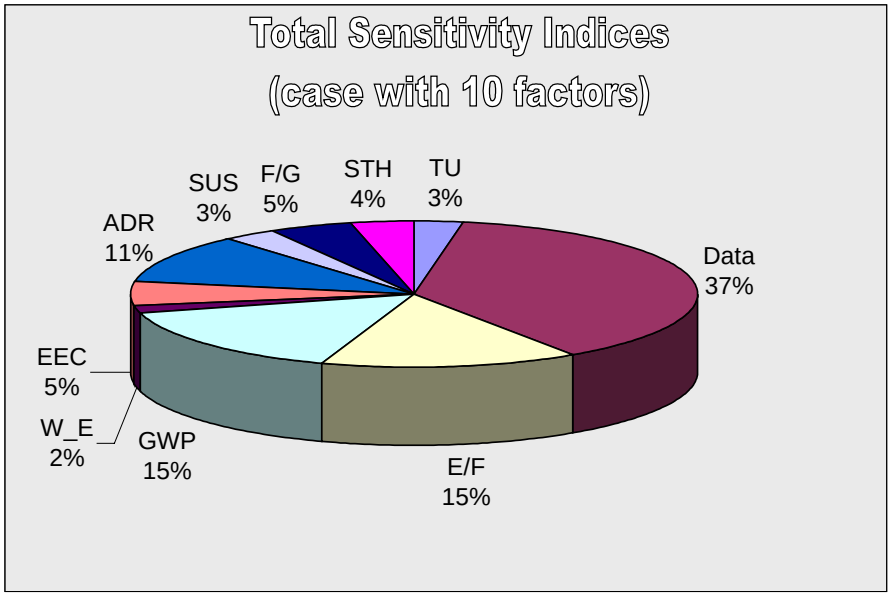


Figure 3.7: Results of SA for the third exercise, where the indicators of eco-toxicological effect (ET) and of Dispersion of Toxics (DT) are included.

Table 5.1: regression statistics of Solow growth

	Ex. Var.	b0	S.E b0	t b0	prob. b0	b1	S.E b1	t b1	prob b1	R-sq.
Model 12	q/A	0.45	0.00	414.51	0.00	0.24	0.00	218.73	0.00	1.00
Model 13	q/A	0.92	0.00	253.19	0.00	-0.63	0.01	-66.69	0.00	0.99
Model 14	ln(q/A_)	-0.73	0.00	-577.94	0.00	0.35	0.00	272.22	0.00	1.00
Model 15	ln(q/A)	-0.04	0.00	-8.83	0.00	-0.92	0.01	-86.45	0.00	1.00
Model 16	d(q/A)	0.00	0.01	0.32	0.75	0.00	0.01	-0.17	0.87	0.00
Model 17	dln(qa)	0.00	0.00	0.44	0.67	0.35	0.01	73.21	0.00	1.00

	Ex. Var.	Adj. R-sq.	S.E. reg.	Sum sq. res	Log lh	DW	Akaike.	Schwarz .	F-stat.	Prob(F-sta
Model 12	q/A	0.99	0.00	0.00	171.49	1.15	-11.30	-11.21	47844.33	0.00
Model 13	q/A	0.99	0.00	0.00	135.94	0.50	-8.93	-8.84	4447.06	0.00
Model 14	ln(q/A_)	0.99	0.00	0.00	170.45	0.95	-10.87	-10.78	74103.94	0.00
Model 15	ln(q/A)	0.99	0.00	0.00	134.94	0.47	-8.58	-8.48	7472.97	0.00
Model 16	d(q/A)	-0.03	0.01	0.00	98.23	1.31	-6.42	-6.32	0.03	0.87
Model 17	dln(qa)	0.99	0.00	0.00	165.63	1.94	-10.91	-10.82	5360.03	0.00

Legend:

Ex. Var. :	Exogenous Variable	Adj. R-sq.	Adjusted R squared
b0:	Expected value of coefficient b0	S.E. reg.	Standard error of regression
S.E. b0:	Standard error of coefficient b0	Sum sq. res	Sum of squared residuals
t b0:	t statistic of coefficient b0	Log lh	Log likelyhood
prob. b0:	2 tailed probability of t statistics	DW	Durbin Watson statistics
b1	Expected value of coefficient b1	Akaike.	Akaike. information criterion
S.E b1	Standard error of coefficient b1	Schwarz .	Schwarz . Information criterion
t b1	t statistic of coefficient b1	F-stat.	F. Statistics
prob b1	2 tailed probability	Prob(F-stat)	2 tailed probability of F statistics
R-sq.	R squared		

Table 5.2: regression statistics of capital labor ratio AR(1) model

Dependent Variable: DLOG(K)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.3605	0.179852	2.004421	0.0548

R-squared	0.087067	Mean dependent var	0.008151
Adjusted R-squared	0.087067	S.D. dependent var	0.039579
S.E. of regression	0.037817	Akaike info criterion	-3.678233
Sum squared resid	0.040044	Schwarz criterion	-3.631085
Log likelihood	54.33439	Durbin-Watson stat	1.772458

Table 5.3: regression statistics of Klein 1

	Ex. Var.	b0	S.E b0	t b0	prob. b0	b1	S.E b1	t b1	prob b1	b2	S.E b2	t b2	prob b2
Model 27	C	11.16	1.75	6.37	0.00	-0.30	0.08	-3.37	0.00	1.32	0.07	18.80	0.00
Model 28	d(C)	0.51	0.31	1.63	0.12	0.42	0.16	2.64	0.01	0.52	0.15	3.24	0.00
Model 29	C	9.08	1.41	6.41	0.00	0.64	0.07	8.15	0.00	0.98	0.01	82.14	0.00

	Ex. Var.	R-sq.	Adj. R-sq.	S.E. reg.	Sum sq. res	Log lh	DW	Akaike.	Schwarz .	F-stat.	Prob(F-stat.)
Model 27	C	0.97	0.97	1.05	19.09	-27.91	2.28	3.09	3.24	295.00	0.00
Model 28	d(C)	0.82	0.79	1.27	27.72	-31.64	1.97	3.46	3.61	38.00	0.00
Model 29	C	0.97	0.96	1.18	25.36	-31.78	1.63	3.31	3.46	293.90	0.00

Legend:

Ex. Var. :	Exogenous Variable	Log lh	Log likelyhood
b*:	Expected value of coefficient b*	DW	Durbin Watson statistics
S.E. b*:	Standard error of coefficient b*	Akaike.	Akaike. information criterion
t b*:	t statistic of coefficient b*	Schwarz .	Schwarz . Information criterion
prob. B*:	2 tailed probability of t statistics	F-stat.	F. Statistics
Adj. R-sq.	Adjusted R squared	Prob(F-stat.)	2 tailed probability of F statistics
S.E. reg.	Standard error of regression		
Sum sq. res.	Sum of squared residuals		

Table 5.4: regression results of prices, private wages and public wages AR(1)

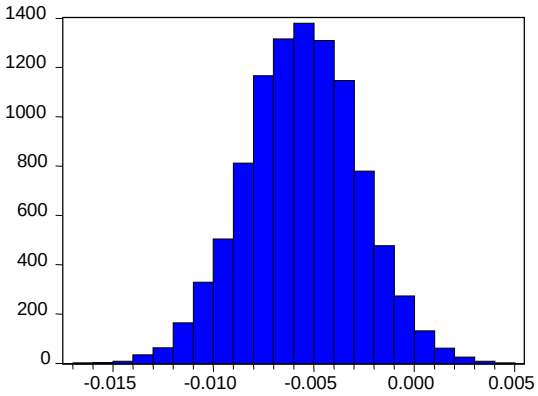
	Ex. Var.	a0	S.E a0	t a0	prob. a0	a1	S.E a1	t a1	prob a1
Model 30	dln(p)	0.02	0.05	0.47	0.64	0.12	0.24	0.51	0.61
Model 31	dln(wp)	0.02	0.02	1.25	0.22	0.24	0.21	1.12	0.27
Model 32	dln(wg)	0.46	0.09	5.13	0.00	-0.63	0.18	-3.47	0.00

	Ex. Var.	R-sq.	Adj. R-sq.	S.E. reg.	Sum sq. res.	Log lh	DW	Akaike.	Schwarz .	F-stat.	Prob(F-st
Model 30	dln(p)	0.02	-0.04	0.23	0.90	2.00	1.86	0.00	0.09	0.27	0.61
Model 31	dln(wp)	0.06	0.01	0.08	0.12	20.82	1.40	-1.98	-1.88	1.26	0.27
Model 32	dln(wg)	0.41	0.37	0.32	1.73	-4.18	1.99	0.65	0.75	12.00	0.00

Legend:

Ex. Var. :	Exogenous Variable	Log lh	Log likelyhood
b*:	Expected value of coefficient b*	DW	Durbin Watson statistics
S.E. b*:	Standard error of coefficient b*	Akaike.	Akaike. information criterion
t b*:	t statistic of coefficient b*	Schwarz .	Schwarz . Information criterion
prob. B*:	2 tailed probability of t statistics	F-stat.	F. Statistics
Adj. R-sq.	Adjusted R squared	Prob(F-stat.)	2 tailed probability of F statistics
S.E. reg.	Standard error of regression		
Sum sq. res.	Sum of squared residuals		

1 Figure 5.1: Probability Density Function of Solow Model Conditional Forecast



Series: Y111	
Sample 1 10000	
Observations 10000	
Mean	-0.005583
Median	-0.005565
Maximum	0.004235
Minimum	-0.016207
Std. Dev.	0.002806
Skewness	-0.019075
Kurtosis	2.997206
Jarque-Bera	0.609682
Probability	0.737240

Figure 5.2: Sensitivity analysis of the variance of conditional forescat of Solow growth model

1.1.1

Parameter	Index	Value
b0	S1	0.0046
b1	S2	0.0007
a1	S3	0.9833
b0,b1	S12	0.0001
b0,a1	S13	0.0001
b1,a1	S23	0.0001

Note: a1 being the parameter of the propagation process

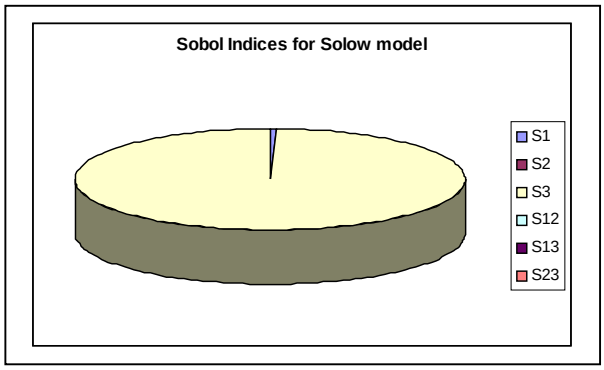
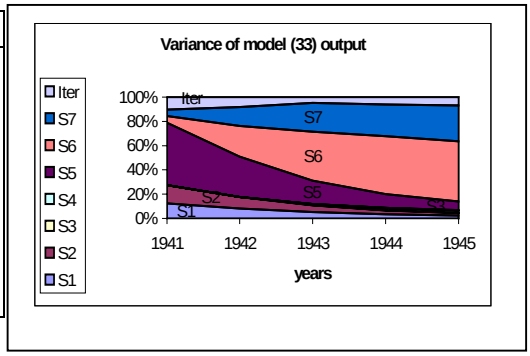


Figure 5.3: Output of Sensitivity Analysis of Klein1 consumption function

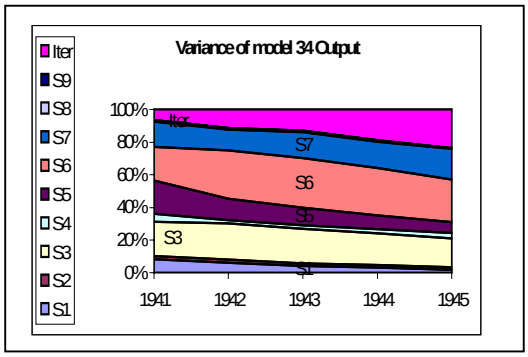
Model (33)

year	1941	1942	1943	1944	1945
S1	0.12	0.08	0.05	0.03	0.02
S2	0.15	0.09	0.05	0.03	0.02
S3	0.00	0.00	0.01	0.02	0.02
S4	0.00	0.00	0.00	0.00	0.00
S5	0.51	0.33	0.19	0.11	0.07
S6	0.05	0.25	0.40	0.47	0.49
S7	0.05	0.15	0.28	0.26	0.29
Iter	0.10	0.08	0.04	0.06	0.07



Model (34)

year	1941	1942	1943	1944	1945
S1	0.08	0.06	0.04	0.03	0.02
S2	0.01	0.01	0.01	0.01	0.01
S3	0.21	0.22	0.21	0.19	0.17
S4	0.04	0.02	0.02	0.02	0.03
S5	0.20	0.13	0.10	0.08	0.06
S6	0.20	0.29	0.30	0.29	0.26
S7	0.15	0.12	0.16	0.16	0.18
S8	0.00	0.00	0.00	0.00	0.00
S9	0.00	0.00	0.00	0.00	0.00
Iter	0.06	0.11	0.13	0.19	0.23



Model (35)

year	1941	1942	1943	1944	1945
S1	0.11	0.07	0.03	0.02	0.01
S2	0.19	0.13	0.08	0.04	0.02
S3	0.00	0.07	0.04	0.03	0.02
S4	0.00	0.07	0.03	0.02	0.01
S5	0.45	0.33	0.21	0.13	0.08
S6	0.19	0.19	0.35	0.43	0.47
S7	0.00	0.10	0.18	0.24	0.27
S8	0.00	0.00	0.00	0.00	0.00
S9	0.00	0.00	0.00	0.00	0.00
Iter	0.02	0.01	0.03	0.05	0.07