Essays on Nonlinear Time Series Analysis and Health Economics

STOCKHOLM SCHOOL
OF ECONOMICS
HANDELSHÖGSKOLAN I STOCKHOLM

EFI, The Economic Research Institute

EFI Mission
EFI, the Economic Research Institute at the Stockholm School of Economics, is a scientific institution that works independently of economic, political and sectional interests. It conducts theoretical and empirical research in the management and economic sciences, including selected related disciplines. The Institute encourages and assists in the publication and distribution of its research findings and is also involved in the doctoral education at the Stockholm School of Economics. At EFI, the researchers select their projects based on the need for theoretical or practical development of a research domain, on their methodological interests, and on the generality of a problem.

Research Organization
The research activities at the Institute are organized into 22 Research Centres.
Centre Directors are professors at the Stockholm School of Economics.

EFI Research Centre:	*Centre Director:*
Management and Organisation (A)	Sven-Erik Sjöstrand
Centre for Entrepreneurship and Business Creation (E)	Carin Holmquist
Public Management (F)	Nils Brunsson
Information Management (I)	Mats Lundeberg
Centre for People and Organization (PMO)	Andreas Werr (acting)
Centre for Innovation and Operations Management (T)	Christer Karlsson
Centre for Risk Research (CFR)	Lennart Sjöberg
Economic Psychology (P)	Guje Sevón
Centre for Consumer Marketing (CCM)	Magnus Söderlund
Centre for Information and Communication Research (CIC)	Per Andersson (acting)
Marketing, Distribution and Industrial Dynamics (D)	Björn Axelsson
Centre for Strategy and Competitiveness (CSC)	Örjan Sölvell
Centre for Business and Economic History (BEH)	Håkan Lindgren
Accounting and Managerial Finance (B)	Johnny Lind
Centre for Financial Analysis and Managerial Economics in Accounting (BFAC)	Kenth Skogsvik
Finance (FI)	Clas Bergström
Centre for Health Economics (CHE)	Bengt Jönsson
International Economics and Geography (IEG)	Mats Lundahl
Economics (S)	Lars Bergman
Economic Statistics (ES)	Anders Westlund
Law (RV)	Erik Nerep
Centre for Tax Law (SR)	Bertil Wiman

Chair of the Board: Professor Carin Holmquist
Director: Associate Professor Filip Wijkström

Address
EFI, Box 6501, SE-113 83 Stockholm, Sweden • Website: www.hhs.se/efi/
Telephone: +46(0)8-736 90 00 • Fax: +46(0)8-31 62 70 • E-mail efi@hhs.se

ESSAYS ON NONLINEAR TIME SERIES ANALYSIS AND HEALTH ECONOMICS

Anna Ovanfors

STOCKHOLM SCHOOL
OF ECONOMICS
HANDELSHÖGSKOLAN I STOCKHOLM

EFI, The Economic Research Institute

Dissertation for the Degree of Doctor of Philosophy, Ph.D.
Stockholm School of Economics, 2006

ISBN nr 91-7258-683-4

Key words:
Non-Linearity, Smooth Transition Autoregression, Common Features, Co-Shifting, Co-Seasonality, Monte Carlo Simulation, Asthma Treatment, Health Related Quality of Life, Patient Satisfaction, Partial Least Squares, Anticoagulant Treatment, Conjoint, Contingent Valuation, Willingness to Pay.

Printed by:
Elanders Gotab, Stockholm 2006

Distributed by:
EFI, The Economic Research Institute
Stockholm School of Economics
P O Box 6501, SE 113 83 Stockholm, Sweden
www.hhs.se/efi

Acknowledgements

In writing this thesis I have benefited greatly from a number of people: colleagues, friends and family. You know who you are and I hope you this way will feel my gratitude towards you.

I thank my advisors: Timo Teräsvirta, Anders Westlund and Magnus Johannesson for your support.

The Research & Development team at AstraZeneca gave me the great opportunity to write the last two chapters.

Without the encouragement and inspiration I have received from Ingela Ternström and Brigitta Hultblad, my parents and my brothers I would definitely not have finished my dissertation. My family: Lennart, Angelinn and Isabelle, have been tremendously supportive and understanding, and without you this work would not have been possible.

Stockholm, 22 May, 2006

Anna Ovanfors

Contents

Introduction

This dissertation consists of four chapters treating two different themes in the field of economic statistics. The first two chapters concern methodological contributions in smooth transition autoregressive (STAR) modelling. The first essay on this theme focuses on modelling single-equation relationships and the second essay develops understanding in methods of modelling systems of equations.

The last two chapters regard contributions to the understanding of applied health economics. The first essay on this theme regards patient satisfaction and how to make optimal improvements within asthma treatment area in Sweden. The second essay compares two different methods of estimating the patient's willingness to pay for improved treatment. This work is applied in the area of anticoagulant treatment in Sweden.

Contents:

Chapter I, *The Net Barter Terms Of Trade: A Smooth Transition Approach.* (Co-authored with Timo Teräsvirta. Published in the **International Journal of Finance and Economics**, Vol. 8, 2003, pp81-97).

The modelling of time series islikely to be drastically affected by the selected starting point of whether or not the series is stationary. The claim of the
Grilli-Yang series on relative prices of primary commodities having a negative linear trend has generated considerable empirical research. This essay, however, examines how much evidence there is against the hitherto neglected assumption that the series is stationary, and our alternative to a linear autoregressive model is the Smooth Transition Autoregressive (STAR) model.

These types of models allow for smooth transitions between a continuum of extreme regimes. The bounded function of the continuous transition variable, which may be a lagged value of the modelled variable or another observable variable, determines the locally linear transitions.

The nonlinearity in the logarithmed series, supported by statistical tests, is described by an Exponential STAR (ESTAR) model. The power of the linearity test against ESTAR when data have been generated by a linear model with a time trend is investigated. The dynamics of the model are illustrated by use of parametrically estimated local spectra and generalized impulse response functions. Our model encompasses models from several previous studies and we demonstrate that a careful study of this time series lends support to the starting point of weak stationarity which is in contrary to the hypothesis of a negative linear trend.

Chapter II, *Co-shifting.*

The purpose of this essay is to develop tests for detecting common features shared by autoregressive processes in a VAR system where the shifts are assumed to be smooth, while so far existing tests have taken only abrupt common trendbreaks into account. Common nonlinear smooth changes in the time-varying intercept of a pair of economic time series are defined as co-

shifting in the essay whereas common nonlinear smooth seasonal patterns are refered to as co-seasonality.

Monte Carlo simulations are used to investigate the empirical size and power of the test statistics that are applied to two time series, Japanese income and Japanese consumption. These series turn out to be weakly and strongly co-shifting according to the performed test.

Chapter III, *Measuring Asthma Patient Satisfaction in Sweden using Partial Least Squares.* (A simpler version of this paper was published in the **International Journal of Health Care Quality Assurance**, Vol. 17, 4, 2004, pp 221-229).

The main purpose of this essay is to identify the most important characteristics involved in overall asthma treatment, with the intention of making powerful recommendations for optimized improvements. Patient perceptions of the separate factors involved in asthma treatment and the impact of each factor on overall satisfaction is investigated using Partial Least Squares (PLS), which are well suited for structural equation modeling when the distribution of the variables is unknown or differs severely from a normal distribution.

The structure of the suggested model is well supported by the data and has been tested on 599 respondents from a questionnaire survey.

The results show that optimizing improvements in overall asthma treatment requires efforts to strive for higher scores for "The physician's manner" and "The physician's medical competence".

The resource use of patients who grade their health-related quality of life high and low respectively is compared. The study implies that improvements in the specific aspect of health-related quality of life lead to reduced resource use.

The dissertation version differ from the published version in the following way:
An estimation algorithm, which we refer to as double PLS-estimation, to reduce remaining multicollinearity in the model is added to this study. Bootstrapping technique is performed to simulate the otherwise unknown precision for the estimated coefficients when using the double PLS-estimation.

Furthermore, the dissertation version also contains some additional variance analysis regarding; cost for patients treated at asthma health care centres versus patients treated at ordinary health care centres, and cost for patients in the different treatment groups.

Chapter IV, *Willingness to pay for anticoagulant treatment: A comparison of conjoint analysis and contingent valuation.* (Accepted for publication in the **PharmacoEconomics**).

Willingness to pay estimates provide information on how patients value different aspects of their overall treatment, and thus they are a worthy tool in the decision of whether the drug should be included in the health insurance. The aim of this study is twofold: to examine preferences and willingness to pay for improved oral anticoagulant treatment, and to compare conjoint analysis and contingent valuation when estimating the patients' mean willingness to pay for alternative treatment.

682 patients, currently treated with warfarin, are asked to value the treatment characteristics corresponding to a new oral drug, of which they have no experience, along with aspects of their current treatment. Medical studies indicate that the effect of the new drug, with fewer side effects and fewer limitations in daily life, is as good as, and in many times better than, the standard warfarin treatment.

To test for a scope effect in the contingent valuation study, the limitations in daily life are varied between "few limitations in daily life" and "many limitations in daily life" in two different sub-samples. In the conjoint analysis, the patients rank 18 treatment concepts including the standard treatment and the two treatment alternatives in the contingent valuation study.

Confidence intervals and variances for mean willingness to pay estimates have been bootstrapped, and thus significant differences in mean willingness to pay estimated using contingent valuation and conjoint analysis can be tested for.

The analysis suggest that patients are willing to pay a substantial amount for the improved anti coagulant treatment, and that the dichotomous choice contingent valuation method and conjoint analysis produce relatively similar results. The estimated scope effect is significantly smaller with contingent valuation than with conjoint analysis.

THE NET BARTER TERMS OF TRADE : A SMOOTH TRANSITION APPROACH

Anna Persson and Timo Teräsvirta
Department of Economic Statistics, Stockholm School of Economics
Box 6501, SE-113 83 Stockholm, Sweden

August 7, 2005

Abstract

This paper analyzes the net barter terms of trade measured by the primary commodity price index relative to the indexes of unit values of export of manufactures from industrial countries. The starting point is that the series is stationary but possibly non-linear. Statistical tests indicate that the logarithmed series is non-linear, and we estimate a Smooth Transition Autoregressive model to describe the process. The dynamics of the model are illustrated by use of parametrically estimated local spectra and generalized impulse response functions. Our model encompasses models from several previous studies, and our conclusion is that the starting point very much decides the outcome.

Key words. Development economics, encompassing, generalized impulse response function, local spectrum, non-linearity, raw material prices, smooth transition autoregression.

JEL Classification Codes. C22, C52, O13

Acknowledgements. Financial support from the Swedish Council for Research in Humanities and Social Sciences is gratefully acknowledged. Earlier versions of the paper have been presented at the European Economic Association Meeting, Berlin, September 1998, and the Nordic Econometric Meeting, Uppsala, May 1999. We wish to thank the participants, Matias Eklöf in particular, for their comments. We are grateful to Matthias Lutz for allowing us to use his time series, Joakim Skalin for his expertise in programming, and a referee for useful remarks. We also wish to thank Pravin Trivedi and Hans Malmsten for helpful discussions and comments. The responsibility for any errors and shortcommings in the paper remains ours.

1. Introduction

The claim of Prebisch (1950) and Singer (1950) that the relative prices of primary commodities have a negative trend, has generated a lot of empirical research. Economically their claim implies that in relative terms, the developed countries gain more and more from the trade with developing countries. Several researchers have since found support for this argument by taking an appropriate price index and fitting a linear trend to it. The coefficient estimate of this trend has turned out to be negative, either for the pre-World War II period (Spraos, 1980) or also for the post-World War II period (Sapsford, 1985). The latter finding was based on assuming a structural shift in the series in 1950. Thirlwall and Bergvin (1985) found negative trends for individual commodity price series for the post-World War II period. Grilli and Yang (1988), who constructed the primary commodity price index for the period 1900-1986, obtained similar results. Some of these studies did, however, ignore the possibility that the residuals of the model, after fitting a constant and a linear trend to the data, are nonstationary. If this is the case, standard inference is not valid and the findings can be questioned.

Cuddington and Urzúa (1989) avoided this difficulty by assuming that the Grilli-Yang series had a unit root. They tested this hypothesis and did not reject it. Discussing the findings of Sapsford (1985) they further argued that "there is no particular reason to assume that a structural shift occurred only in 1950, rather than to assume that such shifts are of regular occurrence - as they would if certain shocks are 'permanent' while others are 'temporary'." Cuddington and Urzúa then assumed a structural shift of their own in 1921 (but not one in 1950 or any other point of time) and fitted a MA(3) model, completed with an impulse dummy to account for the shift, to the difference series. The main reason for introducing this structural shift seems to have been large residual in 1921 in their original model, although a worldwide decrease in wholesale prices in the aftermath of World War I was used as a justification. The main conclusion of the modelling exercise was that the price series was difference stationary without a drift parameter but with a break in 1921. This clearly contradicted the hypothesis of Prebisch and Singer.

Ardeni and Wright (1992) in turn claimed that the hypothesis of difference stationarity was not sufficiently supported by the data. Furthermore, they also argued that the hypothesis of stationarity of the original series had received insufficient attention in the literature. Their solution was to avoid both hypotheses and model the series consisting of the years 1900 to 1988 using a structural time series model; see Harvey (1989, pp.31-51). The results seemed to reconfirm the Prebisch-Singer hypothesis as the trend component turned out to have a deterministic (negative) slope. On the other hand, the authors reported that there was a 64-year cycle in the data. They also pointed out that their model fitted the data better than that of Cuddington and Urzúa. Moreover, Ardeni and Wright noted that the negative trend was unaffected by the assumption of

a structural break in 1921, but, on the other hand, they did not find the evidence for such a break in the series convincing.

Newbold and Vougas (1996) found that the strength of the evidence for the Prebisch-Singer hypothesis depended on whether the time series of relative prices were assumed to be trend stationary or integrated of order one. In the former case, the analysis supported the Prebisch-Singer hypothesis, in the latter it did not. The authors also introduced temporal structural breaks and breaks in the trend in this models but noticed that they did not affect the conclusion.

Trivedi (1995) examined the Grilli-Yang data from 1900 to 1986 by allowing for an unknown structural break in the trend. Helg (1991) also made use of this idea. Trivedi (1995) applied the statistical theory in Zivot and Andrews (1992) and tested the null hypothesis of a unit root against alternatives of stationarity with a break-point, estimated by means of DGP's allowing for change in the mean and the slope of the trend. He was not able to reject the unit root hypothesis. The evidence in the sample pointed at a break-point in 1920. Assuming this break known it was possible to reject the null hypothesis at the 10% significance level. Such a rejection implies accepting trend stationarity and hence finding support for the Prebisch-Singer hypothesis. Following a critical discussion of the data set the conclusion of Trivedi (1995) was that the empirical results are not clear-cut and that more work would be needed to obtain more robust unit root tests than those hitherto applied in this context.

The bivariate study by Powell (1991) argued that all individual price series were integrated of order one and were thus difference stationary. He rejected the Prebisch-Singer hypothesis in favor of structural breaks in 1921, 1937 and 1975.

Lutz (1999) extended the Grilli-Yang series to cover the years 1900-1995. In this study univariate and bivariate time series from recent studies were nested. Contrary to other multivariate studies the results by Lutz (1999) supported the Prebisch-Singer hypothesis. This support was weak when a univariate trend stationary model was used in the analysis.

In this paper our initial assumption is that the original series is stationary. Selecting this starting-point means that the hitherto neglected hypothesis of stationarity will receive attention. The purpose of the paper is to consider the validity of the Prebisch-Singer hypothesis from this angle. Should stationarity hold then, taking the extended series of Lutz (1999), this would imply that the relative prices of primary commodities would have been mean reverting between 1900 and 1995. While accepting stationarity as our working hypothesis we also include the possibility of the series being nonlinear. Because all previous work has been based on the assumption of linearity we begin by testing this hypothesis against a parametric nonlinear model. This model is a variant of the well-known smooth transition autoregressive (STAR) model. The results show that starting from the stationarity hypothesis leads to sensible results. In fact, the STAR model we develop encompasses most of the other models that have been applied to

analyzing the problem. We do not claim that we have found the right model, however. A more prudent, and at the same time more realistic, interpretation of our findings is that they support the conclusion of Newbold and Vougas (1996). These authors argued that the present amount of information in the series allows an investigator to arrive at different conclusions depending on the starting-point of the analyses.

The plan of the paper is as follows. Section 2 describes the data. Section 3 is devoted to the STAR modelling of the series. A set of misspecification tests are conducted for the estimated model. The dynamics of the estimated nonlinear model are illustrated by local spectra and by a generalized impulse response function. In Section 4 we compare our model with previous efforts by using encompassing tests. Section 5 contains the conclusions.

2. The Data

Grilli and Yang (1988) analyzed, among other things, the long run movements in nonfuel primary commodity prices (Grilli and Yang Commodity Price Index, GYCPI) relative to the prices of manufactures (Manufacturing Unit Values, MUV), between 1900 and 1986. The ratio GYCPI/MUV is a measure of purchasing power at international prices of non-fuel primary commodities in terms of traded manufactures. It is used as a measure of the distribution of gains from trade between commodities produced in developing countries and manufactures produced in industrial countries, the net barter terms of trade.

In this study we use the extended Grilli-Yang series from 1900 till 1995. Following others we consider the logarithm of the net barter terms of trade series, denoted LPV (Logarithm Price Values), whose values from 1900 to 1995 are graphed in Figure A.1. The index contains the prices of 54% of all non-fuel commodities in trade between 1977 and 1979, with the world export values for each commodity between 1977 and 1979 as weights. Furthermore 49% of all food products, 83% of all non-food agricultural products, and 45% of all metals are covered.

The time series does not have the same quality in all years since the manufacturing unit values contains missing values for the periods 1914-1920 and 1939-1947 due to the two World Wars. Trivedi (1995) discussed possible consequences of this fact as far as detecting structural breaks was concerned. Grilli and Yang thus constructed the MUV index by interpolation. Another problem might arise when it comes to interpretation of the long run movements in the GYCPI/MUV ratio because the effects of the technical progress on price trends of primary commodities are not the same as those on price trends of manufactures.

3. Modelling the LPV series between 1900 and 1995

3.1. Specification, estimation and evaluation

We begin our analysis by assuming that the LPV series is stationary. This implies that the Prebisch-Singer hypothesis does not hold, but our aim is to find out how much evidence there is in the data against this assumption. We also assume linearity, but because the series is rather irregular with sharp fluctuations we must be prepared to consider the possibility that it cannot be adequately characterized by a linear autoregressive model.

Our alternative to the linear autoregressive model is the Smooth Transition Autoregressive (STAR) model

$$y_t = \boldsymbol{\phi}'\mathbf{w}_t + \boldsymbol{\theta}'\mathbf{w}_t G(\gamma, c; s_t) + u_t, \tag{3.1}$$

where $\mathbf{w}_t = (1, y_{t-1}, \ldots, y_{t-p})'$, $\boldsymbol{\phi} = (\phi_0, \phi_1, \ldots, \phi_p)'$, $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_p)'$, and $\{u_t\} \sim nid(0, \sigma^2)$. Furthermore, G is a transition function with the stationary transition variable s_t. In this paper, following Jansen and Teräsvirta (1996), we first assume that either

$$G(\gamma, c; s_t) = (1 + \exp\{-\gamma(s_t - c)\})^{-1}, \ \gamma > 0,$$

the simple logistic (LSTAR(1)) model or

$$G(\gamma, \mathbf{c}; s_t) = (1 + \exp\{-\gamma(s_t - c_1)(s_t - c_2)\})^{-1}, \ \gamma > 0, \ c_1 \leq c_2,$$

the second-order logistic (LSTAR(2)) model, where $\gamma > 0$ and $c_1 \leq c_2$ are identifying restrictions. These functions are bounded between zero and unity. In a univariate setting, s_t is typically a lag of y_t; see, for example, Teräsvirta (1994). In this case, we want the model to mimic sharp fluctuations of the LPV series around the mean of the process. We should also allow a realization to wander either above or below the mean of the process without a strong tendency to return to the mean. Thus we also consider the first differences $\Delta_k y_{t-1} = y_{t-1} - y_{t-(k+1)}$, $k = 1, \ldots, K$, as possible transition variables.

In order to model y_t (LPV) we apply the modelling sequence proposed in Granger and Teräsvirta (1993, ch. 7) and Teräsvirta (1994). It consists of three stages: specification, estimation, and evaluation. The first step of the specification stage is to test linearity against (3.1). In order to do that we first specify a linear AR(p) model for the LPV series to serve as our null hypothesis. Applying *AIC* (Akaike, 1974) and this leads to choosing $p = 3$ as the maximum lag. The set of possible transition variables is defined to be $S = \{y_{t-1}, y_{t-2}, y_{t-3}; \ \Delta_1 y_{t-1}, \Delta_2 y_{t-1}, \ \Delta_3 y_{t-1}\}$. The LM type linearity tests are described in detail in Teräsvirta (1994), and the results appear in Table A.1. It is seen that linearity is just rejected when $\Delta_3 y_{t-1}$ is the transition variable, and we proceed with the assumption that the series has been generated by a STAR model.

Note that testing separately for each potential transition variable means that we do not control the overall significance level of the testing procedure. As tests here are primarily a model building device this is not our main concern. Possible misspecification is likely to be discovered at either the estimation or the evaluation stage of the modelling cycle.

The next empirical issue, the choice between LSTAR(1) and LSTAR(2), is based on the same short test sequence as in Teräsvirta (1994). The tests clearly suggest an LSTAR(2) model as the p-value p_{F_3} is less than p_{F_4} and p_{F_2}, see Teräsvirta (1994) for motivation for this conclusion and Table A.1 for the results.

A referee remarked that our linearity test may simply detect a misspecification when the true process is trend stationary but the trend is ignored in testing. In order to investigate that possibility we fitted an AR(1) model

$$y_t = \alpha_0 + \alpha_1 t + \beta y_{t-1} + \sigma\varepsilon_t$$

where $\{\varepsilon_t\} \sim$nid$(0, 1)$, to the LPV series and used the estimated model to generate 10000 series with both 100 and 1000 observations. It may be noted that the t-value of $\widehat{\alpha}_1$ was about -3.1. We tested linearity of the 20000 series at the 5% level of significance using an AR(2) model without a trend as the null model. The results can be found in Table A.2. They show that the linearity tests only have weak power when $T = 1000$ and none at all when $T = 100$. As the LPV series contains less than 100 observations, the outcome does not give us reason to believe that our test results are just a consequence of ignoring a linear trend in the process.

The estimation of the parameters of the STAR model turned out to be numerically difficult but the results finally obtained suggested setting $c_1 = c_2$. Instead of imposing this restriction directly, however, we switch to the exponential transition function

$$G(\gamma, c; s_t) = 1 - \exp\left\{-\gamma(s_t - c)^2\right\}, \gamma > 0.$$

which is a close approximation to the LSTAR(2) type logistic one with $c_1 = c_2$. The switch is made for ease of interpretation as the exponential transition function is bounded between zero and unity for all positive values of γ and c. On the other hand, the second-order logistic function is bounded between δ and unity where δ, $0 \leq \delta \leq \frac{1}{2}$, is a function of γ, c_1 and c_2. Note that $\delta = \frac{1}{2}$ for $c_1 = c_2$ and $\delta = 0$ for $c_1 \neq c_2$ when $\gamma \longrightarrow \infty$.

This exponential STAR model is estimated by non-linear least squares. The simplifying restrictions $\phi_j = -\theta_j$, $j = 1, 2, 3$, and $\phi_3 = 1 - \phi_1 - \phi_2$ are supported by the data and consequently imposed. These give the model a "local unit root" when $G = 0$. This happens at a single point, so that having "local unit root behavior" in the model is an event with zero probability. However, the model does display "near unit root behaviour", more of this later.

The estimated model with data from 1900 to 1995 has the form

$$y_t = \underset{(0.087)}{1.21}\, y_{t-1} - \underset{(0.12)}{0.39} y_{t-2} + 0.18 y_{t-3} + (-\underset{(0.087)}{1.21}\, y_{t-1} + \underset{(0.12)}{0.39} y_{t-2} - 0.18 y_{t-3}) \times \left[1 - \exp\left(-\underset{(0.050)}{0.236} \cdot \left(\Delta_3 y_{t-1} - \underset{(0.027)}{0.041}\right)^2 \Big/ \hat{\sigma}^2_{(\Delta_3 y_{t-1})}\right)\right] + \hat{u}_t \tag{3.2}$$

$R^2 = 0.82$ $\quad s = 0.0994$
$LJB = 3.79(0.052)$ $\quad LM - \text{ARCH}(2) = 0.58(0.63)$
$Skewness = -0.28$ $\quad Excess\ kurtosis = 0.44$

where $\hat{\sigma}^2_{(\Delta_3 y_{t-1})}$ is the sample variance of $\Delta_3 y_{t-1}$, used to make the slope parameter γ scale-free, and s the standard deviation of the residuals. The figures in parentheses below the coefficient estimates are standard deviation estimates whereas the ones following the values of test statistics are p-values. The residual variance of (3.2) is 86% of that of the corresponding AR(3) model. Furthermore, LJB is the Lomnicki-Jarque-Bera normality test statistic, and $LM-$ ARCH(2) the LM test statistic of no linear autoregressive conditional heteroskedasticity against at most second order ARCH. Note that the model is quite parsimonious with just four free parameters.

We evaluate model (3.2) further by using misspecification tests constructed for STAR models in Eitrheim and Teräsvirta (1996). The results of tests of no error autocorrelation and parameter constancy in (3.2) can be found in Table A.3. The table also contains the results of testing the hypothesis of no additive non-linearity. The null hypothesis is not rejected for any of these tests at any conventional significance level. We take this as an indication that we have an adequate model for LPV in 1900-1995.

3.2. Interpretation of the dynamics of the estimated model

To consider the dynamic properties of the model we first look at the equilibrium solution of the deterministic counterpart of (3.2). Numerical calculations suggest that there exists an equilibrium $y_\infty = 0$ and that it is unique: all the trajectories we computed converged to that point independent of the starting-values. The equation thus seems to satisfy the necessary condition for stability in Granger and Teräsvirta (1993, p. 11).

The transition function of (3.2) as a function of the transition variable is graphed in Figure A.2.c. Every circle corresponds to an observation. It is seen that the function never attains value unity in the sample. Figure A.2.c shows the behaviour of the transition function over time. Note that the series shows "near unit root behaviour" when $|\Delta_3 y_{t-1}|$ is not large, that is, when the transition function has a low value. Figure A.1 and Figure A.2.a-b show that a large majority of observations belong to this regime. The series then fluctuates without a strong tendency to return to the equilibrium level

even when it has wandered rather far away from it. On the other hand, there also exist large values of the transition function in the sample that correspond to "near white noise behaviour". They cause a sharp return towards the equilibrium (zero) level if the process has strayed away from it. The model is thus capable of generating sharp fluctuations that are present in the series; see Figure A.1. Figure A.2.a shows the value of the transition function: note that the early 1920s, 1921 in particular, but also early 1930s, mid-1970s and late 1980s are periods during which the transition function has obtained a few high values. Comparing this information with Figure A.1 gives an idea of how the model works.

Our estimated ESTAR(1) model also gives us a possibility to discuss the existence of cycles in LPV. Our tool for doing that is the local or sliced spectrum; see Skalin and Teräsvirta (1999) for discussion and another application. It is defined for all local lag polynomials

$$1 - \sum_{j=1}^{p} (\phi_j + \theta_j G(\gamma, c; s_t)) z^j$$

with the roots inside the unit circle as follows:

$$f_{yy}(\omega; s_t) = \frac{1}{2\pi} \left[\left\{ 1 - \sum_{j=1}^{p} (\varphi_i + \theta_i G) e^{-ij\omega} \right\} \left\{ 1 - \sum_{j=1}^{p} (\varphi_i + \theta_i G) e^{ij\omega} \right\} \right]^{-1} \tag{3.3}$$

for $-\pi \leq \omega \leq \pi$. The local spectra are estimated by the autoregressive spectral estimation method (see, for example, Priestley, 1981, p. 601), that is, by replacing φ_i and θ_i by their estimates from (3.2). The local spectra standardized to integrate to unity can be found in Figure A.2.d. Every curve corresponds to an observed value of $s_t = \Delta_3 y_{t-1}$. The dominant feature in the figure is the peak at frequency zero that declines as $G \to 1$. Thus, while there are sharp movements in the series no evidence of cycles seems to be present in the local spectra. Cuddington and Urzúa (1989) decomposed the series into a permanent and a cyclical component using their IMA(1, 3) model with a dummy variable for 1921. Their conclusion was not different from ours as the cyclical component was not oscillatory. Note, however, that the local spectra only give a "local" picture of the dynamics. An estmate of the "global" spectrum may in this case, the model being non-linear, only be obtained by simulation; for discussion see Skalin and Teräsvirta (1999).

The dynamic properties of (3.2) may also be characterized by using the generalized impulse response function as defined in Koop et al. (1996). These authors point out that non-linear models are both history- and shock-dependent. This makes it necessary to generalize the standard impulse response function to account for this fact. The

impulse response is thus defined as an average over both history and the distribution of the shocks, where conditional expectations of the process on its past form the baseline. We randomly draw 100 times with replacement from the vector of residuals from the estimated ESTAR(1) model to produce 100 realizations of the shock for each observed value of the time series used as histories. For the prediction sequence $n = 0, 1, ...20$ we generate 800 replications. The first horizon is generated both with and without the randomly drawn shock, whereas the randomly drawn shocks are used as noise for each horizon thereafter. From these two types of horizon sequences we calculate the means of the 800 replications. The generalized impulse response function is calculated by taking the differences between these two means for each of the 21 horizons (see Koop et al. (1996) for details). We use highest density regions (see Hyndman, 1996) to illustrate the estimated densities graphically. The 50% and 95% highest density regions for $n = 0, 1, ...20$ are graphed in Figure A.3. While the shocks appear quite persistent there is little evidence of asymmetry in the response. This accords with the type (ESTAR) of the non-linear model we have selected and estimated.

4. Encompassing previous models

The conclusion of the preceding section indicate that we have managed to build an adequate univariate time series model for LPV in 1900-1995 starting from the assumption that the original series is stationary. The results (not reported here) turn out to be similar for the shorter sample period (1900-1986) as well.

Nevertheless, as the reason for modelling the series is to investigate the Prebisch-Singer hypothesis we have to test our model against it. This is done by extending the ESTAR model by an intercept and a linear trend implied by this hypothesis, where the residuals are still not autocorrelated. This gives us the Minimal Nesting Model (MNM: see Mizon and Richard, 1986) which nests our specification and the nonstochastic variant of the Prebisch-Singer model with a linear time trend. We then test the joint null hypothesis that the intercept equals zero and the trend variable has a zero coefficient in the MNM. A rejection implies support for the Prebisch-Singer hypothesis. The test is carried out as a standard F-test. From Table A.5.a. it is seen that the null hypothesis cannot be rejected. We may thus conclude that our ESTAR model encompasses (Hendry, 1995, ch. 14) the Prebisch and Singer model, meaning that the structure of the former explains the findings of the latter.

Next we check whether or not our model explains the findings of Cuddington and Urzúa (1989). Since their model contains a unit root and is formulated in first differences we can approximate the MNM by completing model (3.2) linearly by a level dummy variable for 1900-1920 and four unrestricted lags of y_t. The residuals are not autocorrelated. The four lags approximate the IMA(1,3) process in the Cuddington and Urzúa equation (the approximation is quite adequate because the (inverse) roots

of the MA(3) lag polynomial are small in modulus). In this case the encompassing test consists of testing the null hypothesis that the lags and the dummy variable have zero coefficients. The results reported in Table A.5.a indicate that our ESTAR(1) model (3.2) encompasses the Cuddington-Urzúa specification.

Considering the findings of Newbold and Vougas (1996), we approximate the MNM, without autocorrelation in the residuals, by completing model (3.2) linearly by a constant, trend and seven unrestricted lags of y_t. These lags constitute a finite approximation to the AR(∞) representation of the ARMA(2,2) process in Table 9 (higher autoregressive weights are of little importance). The encompassing tests consists of testing the null hypothesis that the constant, the trend and the additional lags have zero coefficients. The results reported in Table A.5.a indicate that our ESTAR(1) model (3.2) encompasses the the ARMA (2,2) model in Table A.13 by Newbold and Vougas (1996).

In Table A.5.b we show the two MNMs when the ability of model (3.2) to encompass the two equations in Trivedi (1995) is considered. The encompassing tests is performed by testing the null hypothesis that the intercept, the dummy, the time trend, lag and lag-differences have zero coefficients in the MNM. From Table A.5.b we can see that our ESTAR model encompasses equations (A.5) and (A.6) in Trivedi (1995).

Building an MNM for comparing our results with those of Ardeni and Wright (1992) appears difficult because the two models have a very different structure. Note, however, that some of their models have a very good fit and they variance encompass model (3.2). The only thing we did was to re-estimate their structural time series model using a more recent version of the STAMP programme, STAMP 5.0. (When we tried to reproduce the results of the authors we noticed that they had used STAMP 3.0. The estimation algorithms in the two versions of the program are different from each other.) Re-estimation changes some of the previous results; see Tables 3 and 4. While STAMP 3.0 yields a deterministic level and slope for the models, these become stochastic when STAMP 5.0 is used for parameter estimation. In the estimation by STAMP 3.0 the cyclical component is reduced to an AR(1) process (no cycle) for all models. Using STAMP 5.0 results in an AR(1) process instead of a cycle for all deterministic trend models but not for stochastic trend models. The slope of the trend is more negative in STAMP 5.0 results. Thus the general tenor of the results remains the same although some details of the models are different from the results obtained by STAMP 3.0. This turns out to be true even for the shorter sample period, from 1900 till 1986.

5. Conclusions

In this paper we find an adequate nonlinear model for the LPV ratio starting from the assumption that the data-generating process is stationary, which excludes any deterministic trend. We also show that our ESTAR model encompasses the Prebisch-Singer

model with a negative linear time trend. It also encompasses the ARMA model based on the assumption of a unit root in the LPV ratio and a structural break in 1921, the breaking trend models of Trivedi (1995) and the "best" specification of Newbold and Vougas (1996). Because the intercept and linear trend model implied by the Prebisch-Singer hypothesis is encompassed by the ESTAR model of this paper, a straightforward conclusion would be that the Prebisch-Singer hypothesis is not valid. Furthermore, the encompassing results suggest that a difference stationary model may not be a viable one either. This, however, may be too hasty an interpretation of the situation. First, Ardeni and Wright demonstrated that at least a stochastic trend variant of the Prebisch-Singer hypothesis is supported by what appears to be a statistical sound model. Although fitting the model to the extended series yields somewhat different results, the general conclusion remains unchanged. Even more importantly, it seems clear that the outcome of the analysis depends on its starting-point. Newbold and Vougas (1996) arrived at the same conclusion. If it is assumed that their variant of the LPV index is trend-stationary, a subsequent modelling exercise leads to accepting the Prebisch-Singer hypothesis. If the starting-point is difference stationarity, that is, a unit root in the price index, then the analysis based on that assumption does not find evidence in favour of this hypothesis. Trivedi (1995) finds that by starting from the assumption of a unit root and a possible trend break at an unknown point the Prebisch-Singer hypothesis does not receive support. If it is assumed that the trend breaks at a known point in the series, in 1920, the conclusion is the opposite. Economically this would mean that developed countries still gains from the trade with developing countries, but not systematically more and more.

Mean reversion in the LPV index is in contrary to the Prebisch-Singer hypothesis. We have demonstrated that if this or, in other words, weak stationarity is selected to be the starting-point of the analysis then a careful study of the time series lends support to this starting-point. Thus the idea that the LPV index is mean reverting, although it may stay far away from its mean extended periods of time, cannot at this stage be dismissed. Our results thus strengthen the conclusion of Newbold and Vougas (1996), namely, that a longer time series is needed to decide whether or not the Prebisch-Singer hypothesis is a feasible one. We could also join Trivedi (1995) and ask for more robust inference. Nevertheless, given the many possibilities for selecting the starting-point, it does not seem likely that the validity or otherwise of the Prebisch-Singer hypothesis is an issue that can be settled with the current amount of sample information.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716-723.

Ardeni, P. G. and B. Wright (1992). The Prebisch-Singer hypothesis: a reappraisal independent of stationarity hypothesis. *Economic Journal,* **102**, 803-812.

Cuddington, J. T. and C. M. Urzúa (1989). Trends and cycles in the net barter terms of trade: a new approach. *Economic Journal*, **99**, 426-442.

Eitrheim, Ø. and T. Teräsvirta (1996). Testing the adequacy of smooth transition autoregressive models. *Journal of Econometrics*, **74**, 59-75.

Granger, C. W. J. and T. Teräsvirta (1993). *Modelling nonlinear economic relationships.* Oxford: Oxford University Press.

Grilli, E. R. and C. Yang (1988). Primary commodity prices, manufactured goods prices, and the terms of trade of developing countries: what the long run shows. *World Bank Economic Review*, **2**, 1-47.

Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman Filter.* Cambridge: Cambridge University Press.

Helg, R. (1991). A note on the stationarity of the primary commodities relative price index. *Economics Letters*, **36**, 55-60.

Hendry, D. F. (1995). *Dynamic econometrics.* Oxford: Oxford University Press.

Hyndman, R. J. (1996). Computing and graphing highest density regions. *American Statistician,* **50**, 120-126.

Jansen, E. S. and T. Teräsvirta (1996). Testing parameter constancy and super exogeneity in econometric equations. *Oxford Bulletin of Economics and Statistics*, **58**, 735-763.

Koop, G., M. H. Pesaran and S. M. Potter (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, **74**, 119-147.

Lutz, M.G. (1999). A general test of the Prebisch-Singer hypothesis. *Review of Development Economics,* **3**, 1999.

Mizon, G. E. and J. -F. Richard (1986). The encompassing principle and its application to non-nested hypothesis tests. *Econometrica*, **54**, 657-678.

Newbold, P. and Vougas, D. (1996). Drift in the relative price of primary commodities: a case where we care about unit roots. *Applied Economics*, **28**, 653-661.

Perron, P. (1989). The great crash, the oil price shock and the unit root hypothesis. *Econometrica,* **57**, 1361-1401.

Perron, P. (1990). Testing for a unit root in a time series with a changing mean. *Journal of Business and Economic Statistics,* **8**, 153-162.

Perron, P. and Vogelsang, T. J. (1992). Testing for a unit root in a time series with a changing mean: corrections and extensions. *Journal of Business and Economic Statistics*, **10**, 467-469.

Powell, A. (1991). Commodity and developing country terms of trade: what does the long run show? *Economic Journal*, **101**, 1485-1496.

Prebisch, R. (1950). *The economic development of Latin America and its principal problems.* New York: United Nations Publications.

Priestley, M. B. (1981). *Spectral analysis and time series.* New York: Academic Press.

Sapsford, D. (1985). The statistical debate on the net barter terms of trade between primary commodities and manufactures: a comment and some additional evidence. *Economic Journal,* **95**, 781-788.

Singer, H. (1950). The distributions of gains between investing and borrowing countries. *American Economic Review,* **40**, 473-485.

Skalin, J. and T. Teräsvirta (1999). Another look at Swedish business cycles, 1861-1988. *Journal of Applied Econometrics* (forthcoming).

Spraos, J. (1980). The statistical debate on the net barter terms of trade between primary commodities and manufactures. *Economic Journal*, **90**, 107-128.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, **89**, 208-217.

Thirlwall, A. P. and J. Bergevin (1985). Trends, cycles, and asymmetries in the terms of trade of primary commodities from developed and less developed countries. *World Development*, **13**, 805-817.

Trivedi, P. K. (1995). Tests of some hypotheses about the time series behavior of commodity process. In: *Advances in Econometrics and Quantitative Economics*:

A Volume in Honor of C. R. Rao, *Eds.* G. S. Maddala, P. C., B. Phillips and T.N. Srinivasan, 382-412, Oxford: Blackwell.

Zivot, E. and W. K. Andrews (1992). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business and Economic Statistics*, **10**, 251-270.

A. Figures and tables

Figure A.1. Logarithmed values of the net barter terms of trade series from 1900 to 1995.

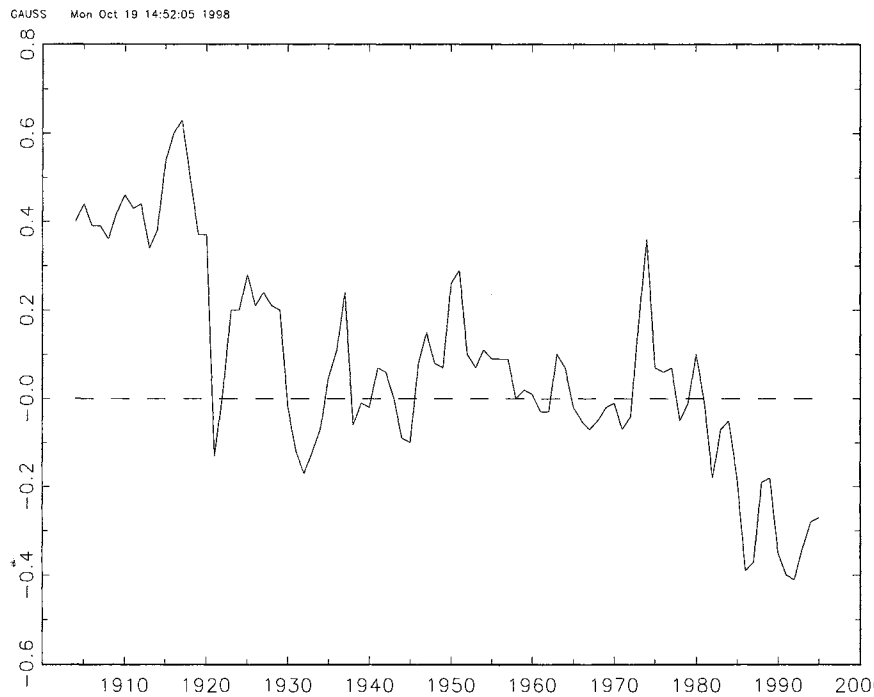
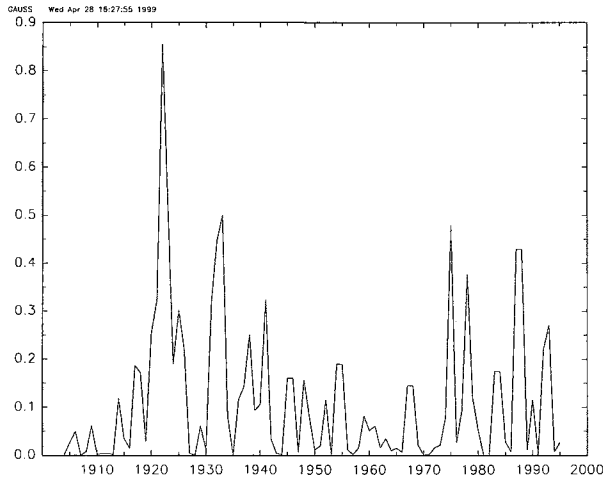


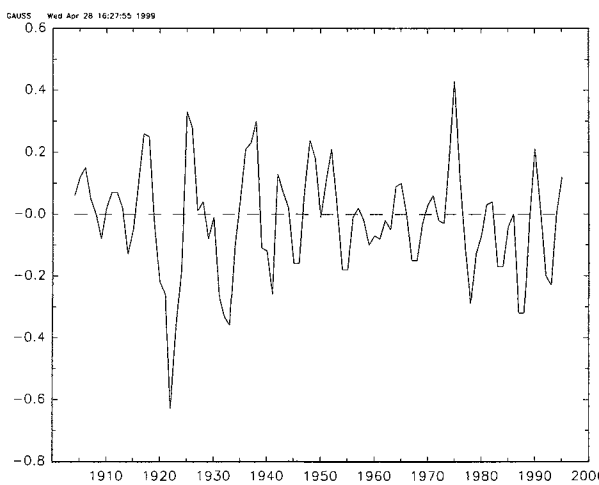
Figure A.2.a-d. Graphs of the transition function and the transition variable

Figure A.2.a.



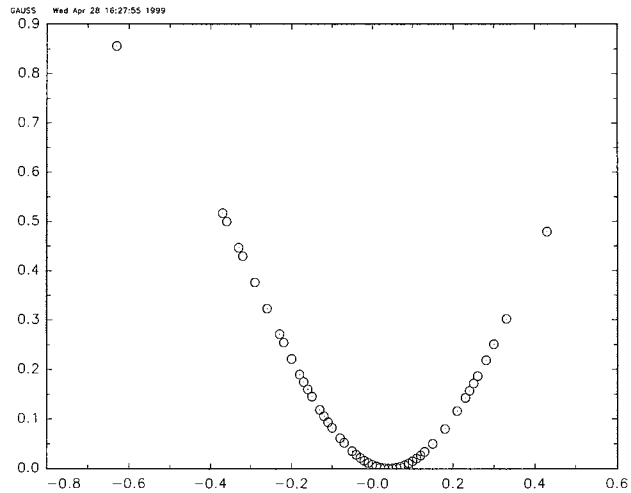
Transition function over time

Figure A.2.b.

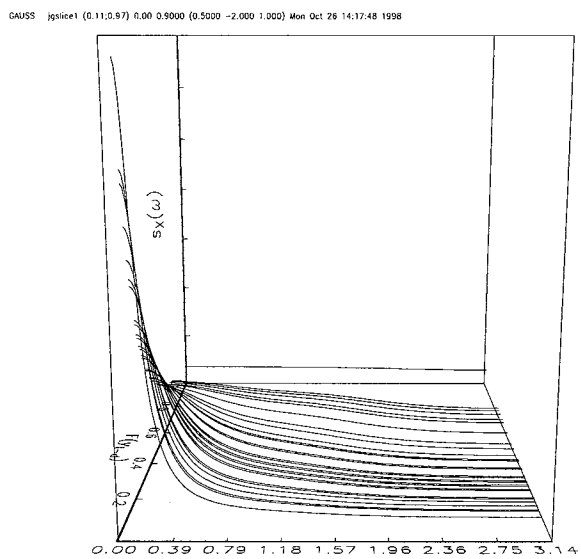


Transition variable over time

Figure A.2.a-d. Graphs of the transition, transition variable and the local spectra
Figure A.2.c.



Transition function as a function
of the transition variable
Figure A.2.d.



Local spectra

Figure A.3. Generalized impuls responses

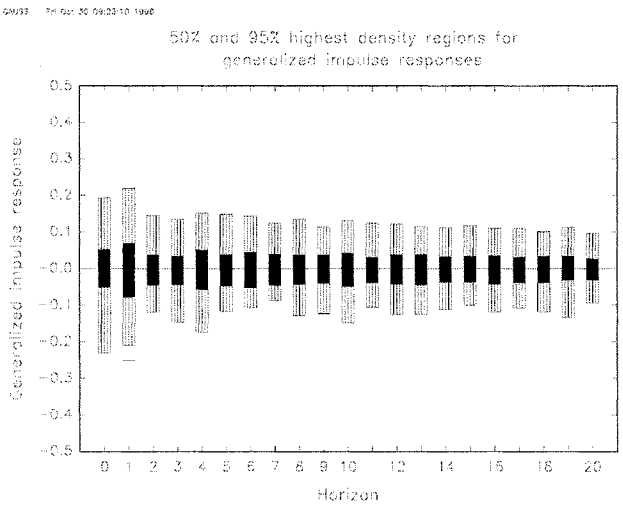


Table A.1. Results of linearity tests.

p-values (p_{F_L}) for different transition variables,
p-values of the tests in the model selection sequence ($p_{F_4}, p_{F_3}, p_{F_2}$),
the selected nonlinear family.

Tests	Transition variable					
	y_{t-1}	y_{t-2}	y_{t-3}	$\Delta_1 y_{t-1}$	$\Delta_2 y_{t-1}$	$\Delta_3 y_{t-1}$
p_{F_L}	0.83	-*)	0.16	0.52	0.35	0.050
p_{F_4}						0.26
p_{F_3}						0.019
p_{F_2}						0.40
STAR						LSTAR(2)

*) Test is not computed. Moment matrix near-singular.

Note: For details on the decision rule in the choice between LSTAR(1) and LSTAR(2) models see Teräsvirta (1994) or Jansen and Teräsvirta (1996).

Table A.2. Power of the linearity test against STAR, when the data have been generated by a linear model with a time trend, for series with 100 and 1000 observations. Nominal significance level is 5%. Number of observations are 10000.

Number of observations	Transition variable					
	y_{t-1}	y_{t-2}	y_{t-3}	$\Delta_1 y_{t-1}$	$\Delta_2 y_{t-1}$	$\Delta_3 y_{t-1}$
100	0.56	0.053	0.043	0.044	0.044	0.041
1000	0.0237	0.233	0.066	0.046	0.096	0.075

Note: The test is the same as the one applied in Table A.1.

Table A.3. Misspecification tests.

p-values of LM tests of no error autocorrelation.

Number of lags			
1	2	3	4
0.46	0.74	0.88	0.79

Note: See Eitrheim and Teräsvirta (1996) for details.

p-values of *LM* type parameter constancy tests.

Parameters	LM_1	LM_2	LM_3
Intercept and lags	0.15	0.11	0.23
Intercept only	0.078	0.19	0.103

Note: LM1 allows for monotonic parameter changes over time. LM2 allows for symmetric parameter changes vis-à-vis a time-point over time.LM3 allow each parameter to change non-symmetrically over time which is the richest alternative of the three. See Eitrheim and Teräsvirta (1996) for details.

p-values of tests of no remaining nolinearity.

Transition variable	p_{F_L}
$s_t = y_{t-1}$	0.55
$s_t = y_{t-2}$	0.19
$s_t = y_{t-3}$	0.72
$s_t = \Delta_1 y_{t-1}$	0.76
$s_t = \Delta_2 y_{t-1}$	0.28
$s_t = \Delta_3 y_{t-1}$	0.35

Note: See Eitrheim and Teräsvirta (1996) for details.

TableA.4. Estimated ESTAR(1) model with the shorter sample between 1900 and 1986

$$y_t = \underset{(0.10)}{1.24} y_{t-1} - 0.24 y_{t-2} + (\underset{(0.10)}{-1.24} y_{t-1} + 0.24 y_{t-2}) \left[1 - \exp\left(\underset{(0.042)}{-0.211} \cdot \left(\Delta_3 y_{t-1} - \underset{(0.022)}{0.061} \right)^2 \Big/ \hat{\sigma}^2_{(\Delta_3 y_{t-1})} \right) \right] + \hat{u}_t,$$

$R^2 = 0.74$ $\quad s = 0.101$

$LJB = 2.81(0.24)$ $\quad LM - \text{ARCH}(2) = 1.62(0.20)$

$Skewness = -0.35$ $\quad Excess\, kurtosis = 0.57$

The residual variance of the above equation is 87% of that of the corresponding AR(2) model.

(A.1)

Table A.5.a Encompassing tests of model (3.2) against other alternatives

Model (3.2) compared to a model implied by the Prebisch-Singer hypothesis. Minimal nesting model (MNM):

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + (1 - \phi_1 - \phi_2) y_{t-3} + \phi_3 + \phi_4 t + (-\phi_1 y_{t-1} - \phi_2 y_{t-2} - (1 - \phi_1 - \phi_2) y_{t-3}) \cdot G(\gamma;\, c;\, \Delta_3 y_{t-1}) + u_t$$

$H_0 : \phi_3 = \phi_4 = 0$

Standard F-test:

$p - value = 0.24$

Model (3.2) compared to Cuddington and Urzúa (1989). MNM:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + (1 - \phi_1 - \phi_2) y_{t-3} + \phi_3 y_{t-3} + \phi_4 y_{t-4} + \phi_5 dum + (-\phi_1 y_{t-1} - \phi_2 y_{t-2} - (1 - \phi_1 - \phi_2) y_{t-3}) \cdot G(\gamma;\, c;\, \Delta_3 y_{t-1}) + u_t$$

$H_0 : \phi_3 = \phi_4 = \phi_5 = 0$

Standard F-test:

$p - value = 0.85$

Model (3.2) compared to the $\underset{ARMA(2,2)}{SBC1}$ model in Table A.9, Newbold and Vougas (1996). MNM:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + (1 - \phi_1 - \phi_2) y_{t-3} + \phi_3 y_{t-4} + \phi_4 y_{t-5} + \phi_5 y_{t-6} + \phi_6 y_{t-7} + \phi_7 + \phi_8 t + (-\phi_1 y_{t-1} - \phi_2 y_{t-2} - (1 - \phi_1 - \phi_2) y_{t-3}) \cdot G(\gamma;\, c;\, \Delta_3 y_{t-1}) + u_t$$

$H_0 : \phi_3 = \phi_4 = \phi_5 = \phi_6 = \phi_7 = \phi_8 = 0$

Standard F-test:

$p - value = 0.61$

Table A.5.b Encompassing tests of model (3.2) against other alternatives

Model (3.2) compared to eq. (A.3), Trivedi (1995). MNM:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + (1 - \phi_1 - \phi_2) y_{t-3}$$
$$+ \phi_0 + \phi_3 DU(\widehat{\lambda})_t + \phi_4 t + \sum_{j=1}^{6} \varphi_j \Delta y_{t-j}$$
$$+ (-\phi_1 y_{t-1} - \phi_2 y_{t-2} - (1 - \phi_1 - \phi_2) y_{t-3}) \cdot G(\gamma;\, c;\, \Delta_3 y_{t-1}) + u_t$$

$2<TB<95$, and $\lambda=\frac{TB}{95}$, $DU(\widehat{\lambda})$ = 1 if t > $\widehat{\lambda}TB$ and 0 otherwise.

$H_0 : \phi_0 = \phi_3 = \phi_4 = \varphi_j = 0 \qquad j = 1...6$

Standard F-test:

$p - value_{TB=1920} = 0.91$ and $p - value_{TB=1946} = 0.62$

Model (3.2) compared to eq. (A.4) in Trivedi (1995). MNM:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + (1 - \phi_1 - \phi_2) y_{t-3}$$
$$+ \phi_0 + \phi_3 DU(\widehat{\lambda})_t + \phi_4 t \phi_3 + \phi_5 D(TB(\widehat{\lambda}))_t + \sum_{j=1}^{6} \varphi_j \Delta y_{t-j}$$
$$+ (-\phi_1 y_{t-1} - \phi_2 y_{t-2} - (1 - \phi_1 - \phi_2) y_{t-3}) \cdot G(\gamma;\, c;\, \Delta_3 y_{t-1}) + u_t$$

We have that $D(TB(\widehat{\lambda}))$ = t-$\widehat{\lambda}TB$ if t > $\widehat{\lambda}TB$ and 0 otherwise.

$H_0 : \phi_0 = \phi_3 = \phi_4 = \phi_5 = \varphi_j = 0 \qquad j = 1...6$

Standard F-test:

$p - value_{TB=1920} = 0.98$ and $p - value_{TB=1946} = 0.71$

A.1. Reestimation of the model of Ardeni and Wright (1992)

A structural time series model with stochastic trend, a cyclical component and a residual component can be written as

$$y_t = \mu_t + \psi_t + \varepsilon_t \quad t = 1, ..., T, \quad \varepsilon_t \sim nid(0, \sigma_\varepsilon^2).$$

where μ_t is the stochastic trend, and ψ_t the cyclical component and ε_t is the residual component. The stochastic trend is defined as

$$\mu_t = \mu_{t-1} + \beta_t + \eta_t \quad t = 1, ..., T, \quad \eta_t \sim nid(0, \sigma_\eta^2),$$

$$\beta_t = \beta_{t-1} + \xi_t \quad t = 1, ..., T, \quad \xi_t \sim nid(0, \sigma_\xi^2).$$

The cyclical component is defined as

$$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \rho \begin{bmatrix} \cos\lambda & \sin\lambda \\ -\sin\lambda & \cos\lambda \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix} \quad 0 \leq \lambda \leq \pi, \quad 0 \leq \rho \leq 1$$

where we assume that κ_t and κ_t^* are uncorrelated irregulars with the same variance σ_κ^2. Furthermore, λ is the cycle frequency in radians, and the damping factor, ρ, of the amplitude is imposed to make the model more flexible. See Harvey (1989) for details and discussion.

When λ equals 0 or π we have the limiting case where the cyclical component reduces to a first-order autoregressive process, with the AR coefficient strictly less than one in absolute value (i.e., constrained to be stationary). The cyclical component is then reformulated as an AR(1) process in the following way

$$\nu_t = \rho_\nu \nu_{t-1} + \kappa_t \quad t = 1, ..., T, \quad \zeta_t \sim nid(0, \sigma_\kappa^2),$$

with ρ_ν in the range $0 < \rho_\nu < 1$.

The disturbances driving the components in the model are mutually uncorrelated.

Ardeni and Wright (1992) estimated two types of models: (a) a stochastic trend model, and (b) a deterministic trend model. When an AR(1) process is proposed by the data it is imposed.

Table A.6. a Structural time series models

Definitions:
σ^2_ε = Irregular, σ^2_η = Level, σ^2_ξ = Slope, σ^2_κ = Cycle, σ^2_p = Pred. error var.,
ρ = Damping factor, λ = Frequency, $(2\pi/\lambda)$ = Cycle period,
L = Log likelihood, R^2_D = Measure of goodness of fit,
$Dum1$: dummy variable defined as 1 in 1921 and 0 elsewhere,
$Dum2$: dummy variable defined as 0 up to 1920 and 1 after that.

Analysis of period 1900-1995 using STAMP *3.0*

Model	σ^2_η	σ^2_ξ	σ^2_κ	σ^2_ε	ρ	Cycle per. $(2\pi/\lambda)$	L	σ^2_p	R^2_D	Slope
(a)	0	0	0.0107	0	0.75	- (AR)	160.43	0.010	0.20	-0.0064
(b)	-	-	0.0107	0	0.75	- (AR)	160.43	0.010	0.20	-0.0063
(a_1)	0	0	0.0091	0	0.79	Dum 1 - (AR)	166.37	0.0084	0.33	-0.0064
(b_1)	-	-	0.0091	0	0.79	- (AR)	166.37	0.0084	0.33	-0.0064
(a_2)	0.0098	0	0	0	0.90	Dum 2 - (AR)	166.15	0.0098	0.22	-0.00042
(b_2)	-	-	0.0089	0	0.82	- (AR)	167.58	0.0082	0.35	-0.0017

able A.6.b Analysis of period 1900-1995 using STAMP *5.0.*

Model	σ^2_η	σ^2_ξ	σ^2_κ	σ^2_ε	ρ	Cycle per. $(2\pi/\lambda)$	L	σ^2_p	R^2_D	Slope
(a)	0.049	0	0.010	0	0.68	9.01	208.41	0.011	0.11	-0.0069
(b)	-	-	0.011	0	0.75	- (AR)	207.43	0.011	0.14	-0.0063
(a_1)	0.0057	0	0.0029	0	0.79	Dum 1 4.92	214.30	0.0091	0.26	-0.0059
(b_1)	-	-	0.0091	0	0.78	- (AR)	212.86	0.0089	0.27	-0.0064
(a_2)	0.0045	0	0.0040	0	0.77	Dum 2 5.74	217.33	0.0085	0.31	-0.0010
(b_2)	-	-	0.0090	0	0.83	- (AR)	214.07	0.0087	0.29	-0.0017

A.2. Reestimation of the model of Trivedi (1995)

First, Trivedi (1995) tested the null hypothesis of a unit root against two alternative models, an additive outlier model and an innovational outlier model as proposed by Perron (1989, 1990) and Perron and Vogelsang (1992). The null model is

$$y_t = \mu + y_{t-1} + \varepsilon_t \tag{A.2}$$

The alternatives to the null are the following *additive outlier model* and *innovational outlier model:*

$$\widetilde{y}_t = \sum_{j=1}^{k} \omega_j D(TB)_{t-j} + \alpha \widetilde{y}_{t-1} + \sum_{j=1}^{k} c_j \Delta \widetilde{y}_{t-j} + \nu_t \qquad t = k+1, ..., T, \tag{A.3}$$

$$\widetilde{y}_t = \mu + \gamma DU_t + dD(TB)_t + \alpha \widetilde{y}_{t-1} + \sum_{j=1}^{k} c_j \Delta \widetilde{y}_{t-j} + \nu_t \qquad t = k+1, ..., T, \tag{A.4}$$

where $\widetilde{y}_t$ denotes the residuals from the regression of y_t on a constant and DU_t, $DU_t = 0$ if $t \leq TB$ and 1 otherwise, $D(TB)_t = 1$ if $t = TB + 1$ and 0 otherwise.

In model (A.4) the mean-shift in the series is not instantaneous so that the level of y depends upon the dynamics of the process. With $TB = 1920$, the t_α-statistic for testing $H_0 : \alpha = 1$ against the alternative of the additive outlier model (A.3) can be found in Table A.1:

Table A.7. t_α-ratios from additive outlier model (A.3)

	k	1	2	3	4	5	6
1900 – 1986	t_α	-4.13**	-3.16	-3.00	-3.43	-2.87	-3.28
1900 – 1995	t_α	-3.49	-2.41	-2.05	-2.21	-1.78	-2.02

** = significant at 5% level. The critical values are -3.68 and -3.40 for the 5 and 10 percent level respectively.

For the shorter sample period the unit root is rejected at the 5 percent level for $k = 1$ but not for other values of k. Due to different value in the Grilli-Yang series in 1945 compared to our series, Trivedi rejected the unit root at the 5 percent level for $k = 1, 2$ but not for larger values of k. For the longer sample period the unit root is not rejected for any k at the 5 percent level.

The values of the t_α-statistic for testing the same null hypothesis against the innovation outlier model (A.4) appear in Table A.2:

Table A.8. t_α*-ratios from innovational outlier model (A:4)*

	k	1	2	3	4	5	6
1900 – 1986	t_α	-4.31**	-3.31	-2.95	-2.91	-2.33	-2.47
1900 – 1995	t_α	-3.55**	-2.43	-1.91	-1.77	-1.30	-1.38

** = significant at 5% level. The critical values are -3.65 and -3.36 for the 5 and 10 percent level respectively.

Table *A.2* says the same about unit root rejection as table *A.1*: for the shorter sample periods the unit root is rejected at the 5 percent level for $k = 1$ but not for larger values of k, and for the longer sample period the unit root is not rejected for any k at the 5 percent level. With Trivedi's series, from 1900 to 1986, the null was not rejected for any values of k.

Second, Trivedi (1995) reexamined the conclusions of Perron and Vogelsang (1992) in accordance with Zivot and Andrews (1992). These authors proposed a new variant of the Perron test in which the break-point is estimated rather than fixed. Here the data-generating processes allow for a change in the mean and the slope of the trend function. The two alternatives to the unit root considered are

$$y_t = \mu + \gamma DU(\widehat{\lambda})_t + \theta t + \alpha y_{t-1} + \sum_{j=1}^{6} c_j \Delta y_{t-j} + \nu_t \tag{A.5}$$

and

$$\widetilde{y}_t = \mu + \gamma DU(\widehat{\lambda})_t + \theta t + \beta D(TB(\widehat{\lambda}))_t + \alpha y_{t-1} + \sum_{j=1}^{6} c_j \Delta y_{t-j} + \nu_t \tag{A.6}$$

where $DU(\widehat{\lambda})_t = 1$ if $t > \widehat{\lambda}T$ and 0 otherwise, $D(TB(\widehat{\lambda}))_t = t - \widehat{\lambda}T$ if $t > \widehat{\lambda}T$ and 0 otherwise. The critical values for $\inf_{\lambda \in \Lambda} t_\alpha(\widehat{\lambda})$ are given in Zivot and Andrews (1992).

To implement the test specifications (A.5) and (A.6) were estimated with the break point TB ranging from $t = 2$ to $t = 87$ for the shorter sample, and to $t = 96$ for the longer sample. The test results can be found in Tables A.3 and A.4.

Table A.9. t_α-ratios from the stationary model with estimated breakpoint, equation (A.5)

	lowest t_α	*second lowest* t_α
1900 – 1986	-4.13 at TB=1920	-2.97 at TB=1949*
1900 – 1995	-3.28 at TB=1920	-3.24 at TB=1945

*) TB=1946 in Trivedi (1995). The 5 percent critical value are -5,08.

Table A.10. t_α-ratios from the stationary model with estimated breakpoint, equation (A.6)

	lowest t_α	*second lowest* t_α
1900 – 1986	-4.13 at TB=1920	-3.49 at TB=1945*
1900 – 1995	-3.54 at TB=1945	-3.24 at TB=1920

*) TB=1946 in Trivedi (1995). The 5 percent critical value are -5,08.

From Tables *A.3* and *A.4* we see that for the case with an unknown breakpoint the null hypothesis of a unit root is not rejected in any case.

Trivedi finds some evidence for a structural break in 1920 using the smaller sample size. This result is time-dependent as the corresponding conclusion from the longer sample 1900 to 1995 is different.

A.3. Reestimation of the model of Newbold and Vougas (1996)

Newbold and Vougas (1996) found that the evidence for the Prebisch-Singer hypothesis depend on whether the time series of relative prices were assumed to be trend stationary or integrated of order one. However, they were using a slightly different series than ours, called GYPI"/MUV (see Grilli and Yang (1988) for details), ranging from 1900 to 1992. In the following we have replicated the analysis by Nwebold and Vougas (1996) for the two sample periods; the shorter sample period from 1900 to1986, and the longer sample period from 1900 to 1995.

We begin by considering the possibility that the generating process is stationary. Let Y_t denote the logarithm of the relative price of primary commodities. Then

$$Y_t = \alpha + \beta t + u_t \tag{A.7}$$

where the random variable u_t is stationary with mean zero. Here we are interested in the slope parameter β. However, it is important also to consider the possibility that u_t is not stationary but rather an autoregressive process with a unit root. Differencing through the above equation gives

$$\Delta Y_t = \beta + v_t \tag{A.8}$$

where the generating process for v_t is stationary and invertible and the parameter of interest is still β.

Newbold and Vougas argue that the statistical distinction between trend stationary and integrated processes will necessarily be very vague. Any model from one class is arbitrarily close to a model from the other. Furthermore they point out that the analysis should be carried out under each of the alternative choices for the order of integration. Our finding is that inference about the parameter β in both the above equations depends rather dramatically on whether the generating model is taken to be trend stationary or integrated of order one.

As a first step, we considered trend-stationary models of the form (A.7), with u_t generated by the ARMA(p, q) process

$$u_t - \phi_1 u_{t-1} - ... - \phi_p u_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1-...-} \theta_q \varepsilon_{t-q} \tag{A.9}$$

where ε_t is white noise. We fitted all possible models with $p + q \leq 6$ to the series. The employed selection criteria are the Schwarz Bayesian Information Criterion (SBIC) and the Akaike Information Criterion (AIC). SBIC is known to estimate consistently the model order when the true model is included in the set considered.

Results from the three highest ranked models according to SBIC (where *SBIC1* is highest ranking) and the preferred model according to AIC can be found in Table A.7.

Table A.11. *Estimated trend-stationary models*

$1900-1986$	$SBIC1$ $ARMA(1,0)$	$SBIC2$ $ARMA(1,1)$	$SBIC3, AIC$ $ARMA(3.3)$
$\widehat{\alpha}$	0.095** (0.038)	0.14** (0.055)	0.0083 (0.015)
$100\widehat{\beta}$	−0.17** (0.69)	−0.25** (0.98)	−0.16 (0.027)
$\widehat{\phi}_1$	0.71 (0.082) **	0.59 (0.13) **	0.83 (1.32)
$\widehat{\phi}_2$			0.53 (2.00)
$\widehat{\phi}_3$			−0.40 (0.72)
$\widehat{\theta}_1$		0.26 (0.52)	−0.094 (1.32)
$\widehat{\theta}_2$			−0.86 (0.41)
$\widehat{\theta}_3$			−0.17 (0.47)

$1900-1995$	$SBC1$ $ARMA(2,2)$	$SBC2$ $ARMA(1,0)$	$SBC3, AIC$ $ARMA(1,1)$
$\widehat{\alpha}$	0.0041 (0.010)	0.097** (0.036)	0.15**(0.053)
$100\widehat{\beta}$	−0.10** (0.018)	−0.18** (0.067)	−0.28** (0.096)
$\widehat{\phi}_1$	1.57** (0.12)	0.72** (0.074)	0.58 (0.12)
$\widehat{\phi}_2$	−0.60** (0.12)		
$\widehat{\theta}_1$	−0.79** (0.14)		0.29 (0.13)
$\widehat{\theta}_2$	−0.32** (0.16)		

** = significance at 5% level. Estimated standard errors within bracets.

A downward drift in relative prices of 0.17% per year in the shorter sample period and 0.10% per year in the longer sample period. The coefficient estimates is significant in the models selected by SBIC1 for both sample periods, and we se no strong evidence for a unit autoregressive root.

Next we estimate ARIMA$(p, 1, q)$ models. Table A.10 summarizes the results in line with Table A.7.

Table A.12. Estimated I(1) models

1900 – 1986	$SBIC1, AIC$ $ARIMA(4,1,2)$	$SBIC2$ $ARIMA(2,1,4)$	$SBIC3$ $ARIMA(1,1,1)$
$100\widehat{\beta}$	−0.26 (0.26)	−0.14 (0.25)	−0.19** (0.087)
$\widehat{\phi}_1$	0.046 (0.47)	0.16 (0.52)	0.70** (0.088)
$\widehat{\phi}_2$	0.29** (0.38)	0.15 (0.48)	
$\widehat{\phi}_3$	−0.027 (0.18)		
$\widehat{\phi}_4$	0.092 (0.13)		
$\widehat{\theta}_1$	−0.41 (0.43)	−0.53 (0.47)	−0.97** (0.019)
$\widehat{\theta}_2$	−0.82 (0.56)	−0.59 (0.65)	
$\widehat{\theta}_3$		−0.047 (0.27)	
$\widehat{\theta}_4$		−0.043 (0.29)	

1900 – 1995	$SBIC1$ $ARIMA(2,1,2)$	$SBIC2$ $ARIMA(1,1,2)$	$SBIC3, AIC$ $ARIMA(3,1,3)$
$100\widehat{\beta}$	−0.19 (0.21)	−0.13 (0.19)	−0.25 (0.029)
$\widehat{\phi}_1$	0.19 (0.31)	0.46** (0.14)	0.17 (0.50)
$\widehat{\phi}_2$	0.22 (0.24)		0.076 (0.39)
$\widehat{\phi}_3$			0.13 (0.25)
$\widehat{\theta}_1$	−0.52** (0.26)	−0.76** (0.14)	−0.46 (0.47)
$\widehat{\theta}_2$	−0.65 (0.34)	−0.36** (0.16)	−0.56 (0.50)
$\widehat{\theta}_3$			−0.18 (0.39)

** = significance at 5% level.

Here the point estimates suggest a downward drift in the relative prices of about 0.2% per year for both sample periods, but the $I(1)$ models provide less evidence of any decline. For both sample periods the point estimate is insignificant for the models selected by SBIC1.

To test the unit root hypothesis formally using the augmented Dickey-Fuller test

we estimated the model

$$\Delta Y_t = \alpha + \beta t + \gamma Y_{t-1} + \sum_{j=1}^{K} \delta_j \Delta Y_{t-j} + \varepsilon_t \qquad \text{(A.10)}$$

for various values of K. The test statistic is the t-ratio associated with $\widehat{\gamma}$ in (A.10). The results are summarized in Table A.9.

Table A.13. t-ratios from augmented Dickey-Fuller regression (equation A.10)

1900 – 1986	K	0	1	2	3	4	5	6
	t_γ	−3.74**	−4.02**	−3.17	−2.97	−2.76	−2.64	−2.36

1900 – 1995	K	0	1	2	3	4	5	6
	t_γ	−3.42**	−3.64**	−2.99	−2.92	−2.73	−2.52	−2.23

** = significance at 5% level.

For both sample periods the null hypothesis of a unit autoregressive root is rejected at the 5% level $K = 0, 1$, whereas it is not rejected for higher values of K. In that case, the approximated autoregression requires K larger than four. The augmented Dickey-Fuller test does not provide strong evidence against the hypothesis that the generating model is integrated of order one.

The conclusion of Newbold and Vougas (1996) holds for our version of the data for both sample periods: while there appears to be evidence of downward drift based on fitted trend-stationary models, the evidence is far less strong when integrated models are estimated. This means that tere are some evidence in favour for the Prebisch-Singer hypothesis if the starting point is that the time series is trend stationary, while if the time series is assumed to be integrated of order on the analysis did not support the Prebisch-Singer hypothesis.

Co-shifting.

Anna Ovanfors
Department of Economic Statistics, Stockholm School of Economics
Box 6501, SE-113 83 Stockholm, Sweden

May 23, 2006

Abstract

This paper focuses on time series with deterministic shifts, and specifically on common features shared by the equations in a VAR system where the shift is assumed to be smooth. We apply the three main test principles in econometrics and an F-approximation for likelihood-ratio based tests for detecting common nonlinear smooth changes in the time-varying intercept of a pair of economic time series *co-shifting*, and common nonlinear smooth seasonal pattern, *co-seasonality*. Monte Carlo simulations are used to evaluate the sample performance of the test statistics. Our tests are applied to two time series, Japanese income and Japanese consumption.

Key words. Common features, VAR system, co-shifting, co-seasonality, monte carlo simulations, Japanese income, Japanese consumption.

1. Introduction

The notion of time series moving together is an important one in econometrics, and the concept of cointegration has a central role in the analysis of nonstationary economic time series. The idea of time series sharing a common feature, for example, the same autocorrelation structure, is a similar but in a way more general concept. It does not require nonstationarity, but the cointegration is a special case of a common feature. The definition of a common feature in two time series or model equations in Engle and Kozicki (1993) is based on linearity. Two series have a common feature if they both have a feature, for example, autocorrelation, but there exists a linear combination of the series that does not have the feature. Engle and Kozicki (1993) tested for a common international business cycle, and they detected a feature by a hypothesis test taking *no feature* as the null, and a *common feature* is detected by a test that finds linear combinations of variables with no feature.

Granger and Teräsvirta (1993) introduced univariate nonlinear models which can account for asymmetries in an economic data series. Anderson and Vahid (1998) applied the results of Granger and Teräsvirta (1993) to multivariate systems of variables. Anderson and Vahid (1998) presented a method of moments test for common nonlinear features in a VAR system. Their test statistics were based on the canonical correlations between the series in the system. Estimating a fully multivariate nonlinear model is complex and it may even be impossible when the sample size is small. Anderson and Vahid (1998) developed a pre-specification test for reducing nonlinearity in a multivariate system. Many economic variables which are interrelated in systems turn out to have common factors even though they are not cointegrated. Earlier studies, like King et al. (1988) and King et al. (1991) and Fama (1992) uses linear specification to investigate such common component models, however Anderson and Vahid (1998) show that a nonlinear specification would be more relevant in these cases.

Instead of entertaining completely flexible nonparametric nonlinear trends and defining the common feature using them, we restrict ourselves to parametric shifts in the processes. These shifts are parameterized using the logistic function as in smooth transition models. More specifically, we assume that the intercept in the individual models we are comparing is time-varying and defined in terms of a logistic transition function. Our aim is to broaden the alternatives of tests to detect common features. In this paper, the feature of the series is that the series is shifting, in other words, the model characterizing the series has a time-varying intercept. More generally, if we consider shifting within a system of equations, a shift means a shift in an equation of the system: the equation in question has a time-varying intercept. The former definition is a special case of the latter in the sense that in the former one, the individual models are univariate and the only links between them are the possibly nonzero contemporaneous error covariances.

For seasonal time series, an extension of the definition is obtained by defining the feature as changing seasonality. In this situation the existence of a linear combination of the series with a constant seasonal pattern is tantamount to co-shifting, which on this occasion is called co-seasonality. This concept is useful in situations where it is of interest to find out whether or not changes in the seasonal pattern of a pair of series have a common source. We develop different test statistics of co-shifting and co-seasonality and we apply them to Japanese disposable income and private consumption.

The plan of the paper is as follows: In section 2 we define co-shifting and co-seasonality, and we discuss models having these co-features. In section 3 we present different test statistics for tests of co-shifting and co-seasonality. Section 4 is a simulation study to investigate the empirical size and power of the tests of co-shifting and co-seasonality. Chapter 5 is devoted to an empirical example and chapter 6 is a conclusion.

2. The models and definitions

Consider the following bivariate vector autoregressive model with a smoothly changing intercept vector:

$$\mathbf{y}_t = \boldsymbol{\lambda}(t) + \boldsymbol{\varepsilon}_t, \ t = 1, \ldots, T \tag{2.1}$$

where $\mathbf{y}_t = (y_{1t}, \ y_{2t})'$, $t = 1, \ldots, T$, are observation vectors, and

$$\boldsymbol{\lambda}(t) = (\lambda_{10k}(t), \ \lambda_{20k}(t))\, \boldsymbol{\lambda}_1 + \mathbf{G}(t^*; \ \boldsymbol{\gamma}, \ \mathbf{c})\boldsymbol{\lambda}_2 \tag{2.2}$$

is the time-varying intercept vector, and $k = 1, 2$, and $t^* = t/T$. Furthermore, $\boldsymbol{\varepsilon}_t \sim \text{IN}(\mathbf{0}, \ \sigma^2\mathbf{I})$. In (2.2) $\boldsymbol{\lambda}_1 = (\lambda_{101}, \ \lambda_{201})'$ and $\boldsymbol{\lambda}_2 = (\lambda_{102}, \ \lambda_{202})'$, whereas the transition matrix

$$\mathbf{G}(t^*; \ \boldsymbol{\gamma}, \ \mathbf{c}) = \text{diag}(G_1(t^*; \ \gamma_1, \ c_1), \ G_2(t^*; \ \gamma_2, \ c_2))$$

$$G(t^*; \ \gamma_j, \ c_j) = (1 + \exp\{-\gamma_j(t^* - c_j)\})^{-1}, \ \gamma_j > 0, \ j = 1, 2 \tag{2.3}$$

equation (2.2) jointly with (2.3), defines a smooth transition in the intercept of model (2.1). The intercept changes smoothly from $\boldsymbol{\lambda}_1$ to $\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2$ over time. If $\gamma_j \to \infty$, then the transition in the intercept becomes a break at c_j. For more discussion on smooth transitions and smooth transition models see, for example, Granger and Teräsvirta (1993) or Teräsvirta (1994).

Using the terminology of Engle and Kozicki (1993), the time-varying intercept of form (2.2) is the feature of the individual equations in model (2.1). We are interested in the case where y_{1t} and y_{2t} evolve in harmony in the sense that the time-varying

intercepts $\lambda_{10k}(t)$ and $\lambda_{20k}(t)$ move together. To this end, we define two co-features for the equations in (2.1):

Definition 1. *Variables y_{1t} and y_{2t} are weakly co-shifting if $\gamma_1 = \gamma_2$ and $c_1 = c_2$ in* (2.1).

Under Definition 1 there exists a linear combination $\boldsymbol{\beta}'\mathbf{y}_t$ of y_{1t} and y_{2t} such that $\boldsymbol{\beta}'\mathbf{G}(t^*; \boldsymbol{\gamma}, \mathbf{c})\boldsymbol{\lambda}_2 \equiv 0$, so that (2.1) collapses into

$$\boldsymbol{\beta}'\mathbf{y}_t = \boldsymbol{\beta}'\boldsymbol{\lambda}(t) + \boldsymbol{\beta}'\boldsymbol{\varepsilon}_t$$

This definition only implies a rather weak relationship between y_{1t} and y_{2t} in the sense that although the variables shift simultaneously, they may shift in different directions. This is the case when λ_{102} and λ_{202} have a different sign. When $\gamma_1 = \gamma_2 \to \infty$, weak co-shifting becomes intercept co-breaking: see, for example, Hendry and Mizon (1998) for a definition. This may be the most common form of co-breaking in practice. The definition of weak co-shifting could be weakened further by substituting $|c_1 - c_2| < \delta$ for $c_1 = c_2$. This would allow for some freedom in the timing of the shift. As discussed above, the definition requires that both equations are nonlinear, that is, the slope parameters, γ_1 and γ_2, are positive. The following definition is stronger than Definition 1 in that the two processes shift in the same direction.

In Bierens (2000) flexible nonparametric nonlinear trends are explored. In this study, two series are nonlinear cotrending if both series are piecewise quadratic and differentiable and those pieces are linearly dependent, i.e. there exist cotrending vectors.

Definition 2. *Variables y_{1t} and y_{2t} are strongly co-shifting if $\gamma_1 = \gamma_2$, $c_1 = c_2$ and $\lambda_{102} = \lambda_{202}$ in* (2.1).

In this case, the "co-feature vector" $\beta = (1, -1)'$ or any nonzero multiple of β. An even stricter definition may be obtained by adding the requirement $\lambda_{101} = \lambda_{201}$ in this definition, but that is not done here. Both definitions are based on the assumption of weak stationarity of $\mathbf{y}_t$. Note that these definitions only make sense if y_{1t} and y_{2t} are measured in the same units. This is usually not a problem if y_{1t} and y_{2t} are logarithmic variables or, for example, interest or unemployment rates. Nevertheless, thaye may be standardized such that they have the same sample standard deviation.

Assume now that $\mathbf{y}_t = \Delta\mathbf{x}_t = (\Delta x_{1t}, \Delta x_{2t})'$ where Δ is a difference operator. Sequence $\{\mathbf{x}_t\}$ is now nonstationary, but we may want to define similar concepts for that case as well. Thus we have the following definitions:

Definition 3. *Variables x_{1t} and x_{2t} are weakly co-bending if $y_t = \Delta x_t$ and $\gamma_1 = \gamma_2$ and $c_1 = c_2$ in* (2.1).

Definition 4. *Variables x_{1t} and x_{2t} are strongly co-bending if $y_t = \Delta x_t$ and $\gamma_1 = \gamma_2$, $c_1 = c_2$ and $\lambda_{102} = \lambda_{202}$ in* (2.1).

Even here, we may allow extra flexibility by substituting $|c_1 - c_2| < \delta$ for $c_1 = c_2$. Note that we do not assume x_{1t} and x_{2t} to be cointegrated. In fact, our definitions

of co-bending are aimed at characterizing some situations where two $I(1)$ variables move together without being linearly cointegrated. Furthermore, in some cases we may assume that $\pi_{ijk} = 0$, $i \neq j$, $k = 1, \ldots, p$.

In general, we have two univariate or bivariate models with a feature: a smoothly time-varying intercept. If the processes are co-shifting or co-bending, then the feature is a co-feature. This means, loosely speaking, that the individual features have common properties. In fact, strong co-shifting and strong co-bending imply that while the individual series or bivariate equations have the feature, there exists a linear combination of them which does not have the feature. We shall return to this property later on.

Consider now a bivariate VAR process with seasonality with a certain periodicity and assume that this seasonality can be adequately characterized by seasonal dummy variables with smoothly time-varying coefficients. Adding these seasonal dummies to (2.1), one obtains

$$\mathbf{y}_t = \mathbf{\Lambda}(t)\mathbf{d}_t + \boldsymbol{\epsilon}_t, \ t = 1, \ldots, T \tag{2.4}$$

where $\mathbf{d}_t = (d_{1t}, \ldots, d_{S-1,t})'$ are the $S-1$ seasonal dummy variables when the period length equals S, and the $(2 \times (S-1))$ smooth transition coefficient matrix

$$\mathbf{\Lambda}(t) = \mathbf{\Lambda}_1 + \mathbf{G}(t^*; \gamma, \mathbf{c})\mathbf{\Lambda}_2 \tag{2.5}$$

with $\mathbf{\Lambda}_k = [\lambda_{ijk}]$, $k = 1, 2$. In model (2.4), the seasonal pattern of the two variables changes smoothly over time according to the transition defined in (2.5). There exists recent univariate evidence of such change in quarterly industrial production series; see Van Dijk et al. (2003). We can consider this change a feature and, analogously to previous considerations, obtain two definitions of a co-feature that is simply called co-seasonality. The first one is as follows:

Definition 5. *Variables y_{1t} and y_{2t} are weakly co-seasonal if $\gamma_1 = \gamma_2$ and $c_1 = c_2$ in (2.4).*

A tighter definition of co-movement in seasonal patterns is also obtained in concordance with previous considerations:

Definition 6. *Variables y_{1t} and y_{2t} are strongly co-seasonal if $\gamma_1 = \gamma_2$, $c_1 = c_2$ and $\lambda_{1j2} = \lambda_{2j2}$, for $j = 0, 1, \ldots, S-1$.*

3. Test statistics

All three classical testing principles, *LM*- *LR*- and *Wald*-test, may be used for deriving an appropriate asymptotic test. The three test statistics have the following definitions:

LM-test:

$$\widehat{lm}/T = tr\left(\left(\widehat{\Omega}_0 - \widehat{\Omega}\right)\widehat{\Omega}_0^{-1}\right) \tag{3.1}$$

LR-test:

$$\widehat{lr} = \left(\left|\widehat{\Omega}\right|\left|\widehat{\Omega}_0\right|^{-1}\right)^{T/2} \tag{3.2}$$

Wald-test:

$$\widehat{w}/T = \mathrm{tr}\left(\left(\widehat{\Omega}_0 - \widehat{\Omega}\right)\widehat{\Omega}^{-1}\right) \tag{3.3}$$

where $\widehat{\Omega}$ and $\widehat{\Omega}_0$ are the residuals from the unrestricted and the restricted systems respectively.

Not that these definitions of the *LM*-test and the *Wald*-test are simple test statistics to compute, as suggested by Doornik and Hendry (1997). They are based upon information from both the unrestricted and the restricted systems, but are easier to compute than the code for the appropriate variance-covariance matrix.

According to Candelon and Lütkepohl (2001) and Doornik and Hendry (1997) an F-approximation for likelihood-ratio based tests might have superior properties, especially in relatively small samples, and the statistic have the following definition:

$$\widehat{Rao} = \frac{1-(1-R_r^2)^{1/s}}{(1-R_r^2)^{1/s}} \cdot \frac{(Ns-q)}{nk_2} \approx F(nk, Ns-q) \tag{3.4}$$

where

$$s = \left(\frac{n^2k_2^2-4}{n^2+k_2^2-5}\right),\ q = \frac{nk_2}{2} - 1,\ N = T - k_1 - k_2 - (n-k_2+1)/2 \tag{3.5}$$

with $k_1$being the number of regressors in the restricted model, k_2 the number of regressors involved in test and:

$$R_r^2 = 1 - \left|\widehat{\Omega}\right|\left|\widehat{\Omega}_0\right|^{-1} \tag{3.6}$$

Se also He et al. (2005) for discussion of the test.

3.1. Testing hypothesis

Definitions 1-4 in section 2 are testable hypotheses within (2.1) and definitions 5-6 within (2.4). Before doing anything else, we have to ensure that the transition functions G_1 and G_2 in (2.2) and in (2.5) are not identically constant, in other words, that the condition $\gamma_j > 0$ holds for $j = 1, 2$. This can be done by testing parameter constancy in both equations of (2.1) and (2.4) respectively by applying the parameter constancy test in Lin and Teräsvirta (1994). If both null hypotheses $\gamma_1 = 0$ and $\gamma_2 = 0$ are rejected, we can consider the present testing problem. We assume that the log-likelihood function of model (2.1) and model (2.4) satisfies the regularity conditions required for consistency and asymptotic normality of the maximum likelihood estimators when $0 < \gamma_j < \infty$, $j = 1, 2$. The null hypothesis of weak co-shifting is

$$H_0 : \gamma_1 = \gamma_2,\ c_1 = c_2 \text{ in (2.1)}. \tag{3.7}$$

The corresponding null hypothesis of strong co-shifting is

$$H_0 : \gamma_1 = \gamma_2,\ c_1 = c_2 \text{ and } \lambda_{102} = \lambda_{202} \text{ in (2.1)}. \tag{3.8}$$

Letting $\gamma^* = \gamma_2 - \gamma_1$ and $c^* = c_2 - c_1$, the null hypothesis of weak co-seasonality is

$$H_0^W : \gamma^* = 0,\ c^* = 0 \text{ in (2.4)} \tag{3.9}$$

The strong co-seasonality is given by the existence of a vector β such that $y_{1t} - \beta y_{2t}$ is linear in (2.4). Under the null of strong co-seasonality we have the following restrictions:

$$\begin{aligned} H_0 : \quad & \gamma_1 = \gamma \\ & \gamma_2 = \gamma \\ & c_1 = c \\ & c_2 = c \\ & \lambda_{1j2} = \beta\lambda_{2j2} \quad j = 0, 1, ..., 3 \end{aligned} \tag{3.10}$$

where β is unknown. The estimation of parameters in the restricted model is carried out using a conditional Nonlinear Least Squares (NLS) procedure and restricted Least Squares (LS). That is, given (β, γ, c), the model is linear and the parameters can be estimated by restricted OLS. In the first step, we assume that (β, γ, c) is given and conditionally on them we estimate the other parameter by restricted LS. In the second step, $(\beta, \gamma, c)'$ is estimated by NLS.

We carried out the restricted LS in a slightly different way. The parameters of the first equation are estimated by LS, and then the parameters $(\lambda_{202}, \lambda_{212}, \lambda_{222}, \lambda_{232})'$ of the second equation are computed using β and the relationship implied by H_0.

Letting $v_t = G(\cdot)\left(\lambda_{202} + \sum_{j=1}^{3}\lambda_{2j2}d_{jt2}\right)$, to estimate $(\lambda_{202}, \lambda_{212}, \lambda_{222}, \lambda_{232})'$ we use the Frisch-Waugh-Lovell Theorem in Davidson and MacKinnon (1993) as follows:

1. Regress y_{2t} on v_t and compute the residuals, say $\widehat{u}$.
2. Regress $x_t = (1, d_{1t2}, d_{2t2}, d_{3t2})'$ on v_t and compute the $(T \times 4)$ matrix X^o which is orthogonal to v.
3. Finally, to estimate $(\lambda_{202}, \lambda_{212}, \lambda_{222}, \lambda_{232})'$ regress $\widehat{u}$ on X^o.

4. Simulation study

In this section we consider the small-sample properties of our test statistics. For weak and strong co-shifting, the simulation procedure starts by generating pair of series which both must be shifting, whereas for weak and strong co-seasonality, we generate pairs of series which both have the kind of seasonality described in section 2. Data generating process for size and power simulations for the weak and strong co-shifting tests:

$$\begin{aligned} y_{1t} &= \lambda_{101} + \lambda_{102}\left(1 + \exp\left(-\gamma_1 \cdot \frac{1}{\sigma^2_{t^*}}(t^* - c_1)\right)\right)^{-1} + \varepsilon_{1t} \\ y_{2t} &= \lambda_{201} + \lambda_{202}\left(1 + \exp\left(-\gamma_2 \cdot \frac{1}{\sigma^2_{t^*}}(t^* - c_2)\right)\right)^{-1} + \varepsilon_{2t} \end{aligned}$$

where $\sigma^2_{t^*}$ is the variance of $t^* = t = T$, used to make the slope parameters γ_1 and γ_1 scale free.

Data-generating process for size and power simulations for the weak and strong co-seasonality tests:

$$\begin{aligned} y_{1t} &= \lambda_{101} + \lambda_{111}d_{1t} + \lambda_{121}d_{2t} + \lambda_{131}d_{3t} + \\ &\quad (\lambda_{102} + \lambda_{112}d_{1t} + \lambda_{122}d_{2t} + \lambda_{132}d_{3t})\left(1 + \exp\left(-\gamma_1 \cdot \frac{1}{\sigma^2_{t^*}}(t^* - c_1)\right)\right)^{-1} + \varepsilon_{1t} \\ y_{2t} &= \lambda_{201} + \lambda_{211}d_{1t} + \lambda_{221}d_{2t} + \lambda_{231}d_{3t} + \\ &\quad (\lambda_{202} + \lambda_{212}d_{1t} + \lambda_{222}d_{2t} + \lambda_{232}d_{3t})\left(1 + \exp\left(-\gamma_2 \cdot \frac{1}{\sigma^2_{t^*}}(t^* - c_2)\right)\right)^{-1} + \varepsilon_{2t} \end{aligned}$$

4.1. Simulation setup for size simulations

The following parameter setup was held fixed for the size simulations:
Weak and strong co-shifting under H_0:

$$\begin{aligned} y_{1t} &= 0.5 + 1.7 \left(1 + \exp\left(-10 \cdot \frac{1}{\sigma^2_{t^*}}(t^* - 0.5)\right)\right)^{-1} + \varepsilon_{1t} \\ y_{2t} &= 0.5 + 1.7 \left(1 + \exp\left(-10 \cdot \frac{1}{\sigma^2_{t^*}}(t^* - 0.5)\right)\right)^{-1} + \varepsilon_{2t} \end{aligned} \tag{4.1}$$

weak and strong co-seasonality under H_0:

$$\begin{aligned} y_{1t} &= 1 + 0.2d_{1t} + 0.3d_{2t} + (-0.4)d_{3t} + \\ &\quad (1 + 0.4d_{1t} + 0.3d_{2t} + (-0.3)d_{3t}) \left(1 + \exp\left(-10 \cdot \frac{1}{\sigma^2_{t^*}}(t^* - 0.5)\right)\right)^{-1} + \varepsilon_{1t} \\ y_{2t} &= 1 + 0.2d_{1t} + 0.3d_{2t} + (-0.4)d_{3t} + \\ &\quad (1 + 0.4d_{1t} + 0.3d_{2t} + (-0.3)d_{3t}) \left(1 + \exp\left(-10 \cdot \frac{1}{\sigma^2_{t^*}}(t^* - 0.5)\right)\right)^{-1} + \varepsilon_{2t} \end{aligned} \tag{4.2}$$

4.2. Simulation setup for power simulations

The following is the parameter setup for the power simulations:
Weak co-shifting under H_1:

$$\begin{aligned} y_{1t} &= 0.5 + 1.7 \left(1 + \exp\left(-10 \cdot \frac{1}{\sigma^2_{t^*}}(t^* - 0.5)\right)\right)^{-1} + \varepsilon_{1t} \\ y_{2t} &= 0.5 + 1.7 \left(1 + \exp\left(-\gamma_2 \cdot \frac{1}{\sigma^2_{t^*}}(t^* - c_2)\right)\right)^{-1} + \varepsilon_{2t} \end{aligned} \tag{4.3}$$

where γ_2 and c_2 were varied in the following way: $\gamma_2 = 15, 20$ and $c_2 = 0.55, 0.58$.

Strong co-shifting under H_1:

$$\begin{aligned} y_{1t} &= 0.5 + 1.7 \left(1 + \exp\left(-10 \cdot \frac{1}{\sigma^2_{t^*}}(t^* - 0.5)\right)\right)^{-1} + \varepsilon_{1t} \\ y_{2t} &= 0.5 + \{1.2\} \left(1 + \exp\left(-\gamma_2 \cdot \frac{1}{\sigma^2_{t^*}}(t^* - c_2)\right)\right)^{-1} + \varepsilon_{2t} \end{aligned} \tag{4.4}$$

where γ_2 and c_2 were varied in the following way: $\gamma_2 = 15, 20$ and $c_2 = 0.55, 0.58$.

Weak co-seasonality under H_1:

$$\begin{aligned} y_{1t} &= 1 + 0.2d_{1t} + 0.3d_{2t} + (-0.4)d_{3t} + \\ &\quad (1 + 0.4d_{1t} + 0.3d_{2t} + (-0.3)d_{3t}) \left(1 + \exp\left(-10 \cdot \frac{1}{\sigma^2_{t^*}}(t^* - 0.5)\right)\right)^{-1} + \varepsilon_{1t} \\ y_{2t} &= 1 + 0.2d_{1t} + 0.3d_{2t} + (-0.4)d_{3t} + \\ &\quad (1 + 0.4d_{1t} + 0.3d_{2t} + (-0.3)d_{3t}) \left(1 + \exp\left(-\gamma_2 \cdot \frac{1}{\sigma^2_{t^*}}(t^* - c_2)\right)\right)^{-1} + \varepsilon_{2t} \end{aligned} \tag{4.5}$$

where γ_2 and c_2 were varied in the following way: $\gamma_2 = 10.7, 11.1$ and $c_2 = 0.508, 0.512$. To get power in the simulation, only small variations in γ_2 and c_2 were needed.

Strong co-seasonality under H_1:

$$\begin{aligned} y_{1t} &= 1 + 0.2d_{1t} + 0.3d_{2t} + (-0.4)d_{3t} + \\ &\quad (1 + 0.4d_{1t} + 0.3d_{2t} + (-0.3)d_{3t}) \left(1 + \exp\left(-10 \cdot \frac{1}{\sigma^2_{t^*}}(t^* - 0.5)\right)\right)^{-1} + \varepsilon_{1t} \\ y_{2t} &= 1 + 0.2d_{1t} + 0.3d_{2t} + (-0.4)d_{3t} + \\ &\quad \{3.5\} + \{1.6\}\, d_{1t} + \{1.5\}\, d_{2t} + \{-1.5\}\, d_{3t} \left(1 + \exp\left(-\gamma_2 \cdot \frac{1}{\sigma^2_{t^*}}(t - c_2)\right)\right)^{-1} + \varepsilon_{2t} \end{aligned} \tag{4.6}$$

where γ_2 and c_2 were varied in the following way: $\gamma_2 = 100, 1000$ and $c_2 = 0.65, 0.8$. Note that to get power in the simulation, were much larger variations in γ_2 and c_2 were needed compared to the weak co-seasonality simulations.

In order to reduce the importance of the starting values, the first 200 observations are excluded from the reported results. That is, $T+200$ observations are generated and the sample sizes used are $T = 200$, 500, 1000.

Using

$$\frac{\widehat{P} - P_{H_0}}{\sqrt{\frac{P_{H_0}(1-P_{H_0})}{n}}} \sim N(0,1), \tag{4.7}$$

with total $N = 1000$ replications reported from each experiment, and $n = 1000$, a 95% confidence interval of [0.0362, 0.0638] can be calculated, which we think is sufficient precision. We obtain a 95% confidence interval of [0.0305, 0.0695] and [0.0192, 0.0808] for $n = 500$ and $n = 200$ respectively.

In order to find sufficient power, $\gamma_2 = 10$ will be generated for the co-shifting and the co-seasonality tests, which gives us a transition function with a typical S-shape as in Figure B.1.

For different simulated numbers of observations, n, different amounts of the trials failed in the simulation exercise:

$n = 200$	$n = 500$	$n = 1000$
93	19	2

4.3. Graphical techniques for reporting result

Davidson and MacKinnon (1997) discuss simple graphical methods that appear to be very useful for characterizing both the size and the power of test statistics, since they convey much more information than tables can do.

The empirical distribution function, or EDF, of the P values of the test statistics is used to create all of the following graphs. In a Monte Carlo experiment N realizations of some test statistics τ are generated using a data-generating process (DGP) that is a special case of the null hypothesis. Denote these simulated values by $\varkappa_j$, $j = 1, ..., N$. The P value of $\varkappa_j$ is the probability of observing a value of $\varkappa$ as or more extreme than $\varkappa_j$, according to some distribution $F(\varkappa)$.

EDF of the p_j's, evaluated at m points q_i, $i = 1, ..., m$, is defined as

$$\widehat{F}(q_i) \equiv \frac{1}{N}\sum_{j=1}^{N} I(p_j \leq q_i) \tag{4.8}$$

where $I(p_j \leq q_i)$ is an indicator function that takes the value 1 if its argument is true and 0 otherwise.

The following choice of the x_i's is recommended in Davidson and MacKinnon (1997)

$$q_i = 0.001, 0.002, ..., 0.010, 0.015, ..., 0.990, 0.991, ..., 0.999 \qquad (m = 215).$$

The P **value plot** is simply a plot of $\widehat{F}(q_i)$ against q_i. If the distribution of τ used to compute the p_j's is correct, the resulting graph should be close to the 45° line. The P **value discrepancy plots,** a plot of $\widehat{F}(q_i) - q_i$ against q_i, is much more informative when we are dealing with test statistics that are "well-behaved". However, some of this information is spurious according to experimental randomness. Davidson and MacKinnon (1997) therefore discuss semiparametrical methods for smoothing them.

In order to plot power against true size, **size power curves**, according to Davidson and MacKinnon (1997), we need to perform two experiments. In the first experiment, the null hypothesis holds, and in the second it does not. Let the points on the two approximate EDFs be denoted $\widehat{F}(q)$ and $\widehat{F}^*(q)$, respectively. Plotting the points $(\widehat{F}(q), \widehat{F}^*(q))$, including the points (0, 0) and (1, 1), generates a size-power curve on a correct size-adjusted basis.

In the next section, we will discuss these types of plots for the simulations.

4.4. Simulation results

4.4.1. The empirical size of the tests

In Appendix A, Table A.1, we can see the results for the size simulations. For weak co-shifting, strong co-shifting and weak co-seasonality simulations all four tests have the correct size when for all n. However, for strong co-seasonality simulations when $n = 200$ the *Wald*-test is significantly slightly oversized when the nominal size is 5%.

In Appendix B the P value plots in Figures B.3 - B.16 contain more information about the size of the tests. We have chosen to show P value plots only for sample sizes 1000 and 200.

Sometimes spurious information in the curves makes them hard to interpret. In such situations Davidson and MacKinnon (1997) suggest smoothing the P value discrepancy plots. However, in our cases the P value discrepancy plots are easy to interpret which is why we have chosen not to present the smoothed P value discrepancy plots.

Firstly, we can observe that the *LM*-, *LR*- and the *Wald*-tests behave very similar in all simulations, while the *Rao*-test behave differenctly and this difference is more obvious in the co-shifting simulations. In Figure B.3 for weak co-shifting when $n = 1000$, we see that the *LM*-, *LR*- and the *Wald*-tests seams to be correctly sized up to the 10%-level, and then they becomes slightly over-sized. The *Rao*-test becomes systematically more and more under-sized for higher size levels. In figure B.4 for weak co-shifting when

$n = 200$ we se that the LM-, LR- and the $Wald$-tests seams to be correctly sized up to the 5%-level, whereafter they then becomes slightly over-sized. The Rao-test becomes systematically more and more under-sized for higher size levels. Figures B.6 and B.7 concern the P value plots for strong co-shifting simulations, and these graphs show the same result as the weak co-shifting simulations.

The P value plots for the weak co-seasonality simulations are shown in Figures B.11 and B.12. For $n = 1000$ we can see that all four tests seems to be correctly sized for all size levels. For $n = 200$ the Rao-test seams to have correct size up to the 18%-level while it then becomes under-rejecting. the LM-, LR- and the $Wald$-tests seams to become over-rejecting for size levels over 2%.

Finally, for strong co-seasonality simulations in Figures B.15 and B.16, the P value plots show that the tests fluctuate around the zero-line with a tendency to be over-rejecting all the time except for the *Rao*-test which have a tendecy to become under-rejecting for size levels over 12% when $n = 1000$. When $n = 500$ we can see that the LM-, LR- and the $Wald$-tests are over-rejecting up to the 10%-level, while they then have a tendency to become under-rejecting. The Rao-test are under-rejecting for all size-levels.

4.4.2. The empirical power of the tests

Appendix A Tables A.2 - A.5 contain the results for the power simulations. We can see that we are able to get reasonable power in all simulation studies for $n = 1000$ and $n = 500$. However, in order to find sufficient power in the strong co-seasonality simulation, the slope parameter in the second equation in model (4.6) is quite large, $\gamma_2 = \{100, 1000\}$, whereas γ_2 lies between 10.7 and 20 in the other simulation studies.

In the case of strong co-seasonality simulations, the transition is rather quick, see Figure B.1 for the transition function as a function of the transition variable, t, of an $LSTAR(1)$ model for different speed of transition.

The size-power curves proposed by Davidson and MacKinnon (1997) in Figures B.5 - B.14 are generated in a way that gives us graphs of the power on a correct, size-adjusted basis. Size-power curves on a correct, size-adjusted basis make conservative tests and over-sized tests hard to interpret. As we can see from the size-power graphs, for the weak and strong co-shifting simulations the *Rao*-test does not reach power as fast as the LM-, LR- and the $Wald$-tests. This is due to the fact that the *Rao*-test, as we have seen from the p value plots, have a tendency to be under-rejecting. For the weak and strong co-seasonality simulations however, all four tests lies so close that they are hard to distinguish from each other.

5. Empirical example

Let's apply the tests to two time series, Japanese income and Japanese consumption, see Figure B.2. We begin by testing for weak and strong co-bending. We transform the series to percentage changes by differencing the logarithms of the original series. The transformation makes them stationary. In order to perform the tests for weak and strong co-bending, we must check if the series are bending. Before doing anything else, we have to ensure that the transition functions are not identically constant, in other words, that the condition $\gamma > 0$ holds for both Japanese income and consumption. The first step is to specify linear AR(p) models for the series. Applying AIC (Akaike, 1974) leads to choosing $p = 10$ for both income and consumption. Next we test the constancy of the intercept in both series (see Lin and Teräsvirta, 1994). The results appear in table A.6. The test shows significant shifting of four-quarter difference for Japanese income and consumption, and changing seasonality of first difference of income and consumption.

We now want to see if the fourth quarter difference of the income and consumption series is co-shifting, weakly and strongly, and if the first difference of japanese income and consumption is co-seasonal, weakly and strongly. We apply the tests proposed in Sections 3 and 4, and the first step is to model the series under the null hypothesis of co-features and under the alternative hypothesis of no co-features. The algorithms do not converge when the models are estimated simultaneously. However, the $Wald$-test we have used in the simulation study is the simpler test, suggested by Doornik and Hendry (1997), that needs information from both the restricted and the unrestricted systems respectively. An ordinary $Wald$-test needs estimation only under H_1 and can easily be derived from the $Wald$-test used in the simulation study:

the log-likelihood is given by:

$$l = c - \frac{T}{2}\ln|\sum| - \frac{1}{2}\sum_{t=1}^{T}(y_t - f(x_t;\theta))'\sum^{-1}(y_t - f(x_t;\theta))$$

where

$$y_t - f(x_t;\theta) = \begin{bmatrix} y_{1t} f_1(x_{1t};\theta_1) \\ y_{2t} f_2(x_{2t};\theta_2) \end{bmatrix}$$

and

$$\sum = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

Furthermore

$$\begin{aligned}\theta_1 &= (\beta'_{11}, \beta'_{12}, \gamma_1, c_1)' \\ \theta_2 &= (\beta'_{21}, \beta'_{22}, \gamma_2, c_2)'\end{aligned}$$

For weak co-shifting we have that $H_0 : \gamma_1 = \gamma_2, c_1 = c_2$. Let $\theta = (\theta'_1, \theta'_2)'$, then $H_0 : R\theta = 0$, where

$$R = \begin{bmatrix} 0' & 0' & 1 & 0' & 0' & 0' & -1 & 0' \\ 0' & 0' & 0' & 1 & 0' & 0' & 0' & -1 \end{bmatrix}$$

The $Wald$-statistic becomes

$$Wald = \widehat{\theta}' R' \left(R\widehat{V}^{-1} R' \right)^{-1} R\widehat{\theta}$$

where $V = cov\left(\sqrt{T}\widehat{\theta}\right)$. $\widehat{V}^{-1}$ is obtained from the estimation under H_1. The $Wald$-test for strong co-shifting, weak and strong co-seasonality respectively can be derived similarly.

The results when we use this $Wald$-test appear in Table A.7. It can be seen that the $Wald$-test indicates that Japanese income and consumption are weakly and strongly co-shifting. The estimated model for co-shifting and co-seasonality under H_1 are shown in Tables A.8 and A.9 respectively.

6. Conclusions

It is important to understand the economic relationship between nonlinear components in time series. Existing tests to detect common features shared by autoregressive processes in a system with trend breaks allow only for an abrupt change.

Vector Autoregression (VAR) systems have proven to be useful for analyzing common features in data series. Engle and Kozicki (1993) introduced a class of statistical tests for the hypothesis that some feature that is present in each of several variables is common to them. In the study by Engle et al. (1993), Japanese disposable income and private consumption are examined. The result of this study is that the support for cointegration at the annual frequency is weak. Furthermore, Hall et al. (1997) tested the sensitivity of the model of Engle et al. (1993) to changes in the sample period using the approach by Hylleberg et al. (1990). The result is that the modelling exercise of Engle et al. (1993) is not robust over time. Hall et al. (1997) instead find that a model allowing for regime changes is supported by the data.

Bierens (2000) recently underlined the need to have a form of co-movement between the series other than cointegration. He defined a number of series as cotrending if they individually are stationary around a nonlinear trend but there exists a linear combination of the series that is trend stationary around a linear trend or a constant. Bierens developed a nonparametric nonlinear trend that allows the nonlinear trend to be very flexible if necessary. On the other hand, trends with discontinuities such as kinked trends also fit into his definition. There is also a useful discussion in the article about the need for a concept of nonlinear cotrending and its possible uses.

In a recent article, Anderson and Vahid (1998) considered nonlinearity as the feature of the time series. If there exists a linear combination of the series that eliminates this nonlinearity, the nonlinearity is a common feature called common nonlinearity. As a possible specification of the feature, the authors used the smooth transition regression model; see, for example, Granger and Teräsvirta (1993) and Teräsvirta (1998). Our approach combines that of Bierens (2000) and Anderson and Vahid (1998). We let the common feature shared by the models of I(0) processes in a VAR system be that the shifting is assumed to adjust smoothly between different values like in Leybourne et al. (1996). Tests are developed for detecting common nonlinear smooth changes in the time-varying intercept of a pair of economic time series, weak and strong *co-shifting*, and common nonlinear smooth seasonal pattern, weak and strong *co-seasonality*. The sample performance of the test statistics of the three main test principles in econometrics and an F-approximation for likelihood-ratio based tests (see for example Doornik and Hendry,1997; and Candelon and Lütkepohl, 2001) is evaluated using Monte Carlo simulations.

From the results for the size simulations for weak co-shifting, strong co-shifting and weak co-seasonality, all four tests have approximately the correct size when the number

of observations in the series, n, is around 1000 and 500. When $n = 200$ the $Wald$-test for the strong co-seasonality simulation is slightly over sized.

We also show that we are able to get reasonable power in all simulation studies when $n = 1000$. When $n = 500$ we have acceptable power in the simulation studies for weak and strong co-shifting but not for weak and strong co-seasonality, and when $n = 200$ we do not have acceptable power in any of the simulation studies. In the weak co-seasonality simulations the Rao-test seams to have better performance than the other three tests, in all other simulation studies the LM-, LR- and the $Wald$-tests behave better.

We apply our tests to the two time series, Japanese income and Japanese consumption, and it can be seen that Japanese income and consumption are weakly and strongly co-shifting according to the $Wald$-test.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716-723.

Anderson, H. M. and F. Vahid (1998). Testing multiple equation systems for common nonlinear components. *Journal of Econometrics*, **84**, 1-36.

Bierens, H. J. (2000). Nonparametric nonlinear cotrending analysis, with an application to interest and inflation in the United States. *Journal of Business and Economic Statistics*, 18, 323-337.

Candelon, B. and H. Lütkepohl (2001). On the reliability of chow-type tests of parameter constancy in multivariate dynamic models. *Economics Letters.* **73**, 155-160.

Davidson, R. and J. G. MacKinnon (1993). *Estimation and inference in econometrics.* Oxford University Press.

Davidson, R. and J. G. MacKinnon (1997). Graphical methods for investing the size and power of hypothesis test. Working paper.

Doornik, J. A. and D. F. Hendry (1997). *Modelling Dynamic Systems Using PcFiml 9.0 for Windows.* International Thomson Business Press.

Engle, R. F., C. W. J. Granger, S. Hylleberg and H. S. Lee (1993). Seasonal Cointegration; the Japanese Consumption Function. *Journal of Econometrics*, **55**, 275-298.

Engle, R. F. and S. Kozicki (1993). Testing for common features. *Journal of Business and Economic Statistics,* **11**, 369-395.

Granger, C. W. J. and T. Teräsvirta (1993). *Modelling nonlinear economic relationships.* Oxford: Oxford University Press.

Hall, S. G., Z. Psaradakis and M. Sola (1997). Cointegration and Changes in Regime; the Japanese Consumption Function. *Journal of Applied Econometrics*, **12**, 151-168.

He, C., T. Teräsvirta and A. Gonzalez (2005). Testing parameter constancy in stationary vector autoregressive models against continuous change. Stockholm School of Economics. SSE/EFI Working Paper Series in Economics and Finance.

Hylleberg, S., R. F. Engle, C. W. J. Granger and B. S. Yoo (1990). Seasonal intergration and Cointegration. *Journal of Econometrics*, 44, 215-238; reprinted in S. Hylleberg (ed), *Modelling Seasonality*, Oxford University Press, Oxford, 1992.

Mizon, G. E. and D. F. Hendry (1998). Exogeneity, causality, and co-breaking in economic policy analysis of a small econometric model of money in UK. *Empirical Economics,* **23**, 267-294.

Leybourne, S., P. Newbold and D. Vougas (1998). Unit roots and smooth transitions. *Journal of Time Series Analysis*, **19**, 83-97.

Lin, C.- F. J. and T. Teräsvirta (1994). Testing the constancy of regression parameters against continuous structural change. *Journal of Econometrics*, **62**, 211-228.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, **89**, 208-217.

Teräsvirta, T. (1998), *Modeling economic relationships with smooth transition regressions*, in: A. Ullah and D.E.A. Giles, eds. Handbook of applied economic statistics, New York: Dekker, 507-552.

Van Dijk, D., B. Strikholm and T. Teräsvirta (2003). The effects of institutional and technological change and business cycle fluctuations on seasonal patterns in quarterly industrial production series. *Econometrics Journal*, **6**, 79-98.

Zivot, E. and D. W. K. Andrews (1992). Further evidence on the great crash, the oil price shock and the unit root hypothesis. *Journal of Business and Economic Statistics*, **10**, 251-270.

A.Tables

Table A.1. Size of the weak and strong co-shifting and the weak and strong co-seasonality tests. Rejection frequencies at the 5%-level of the null hypothesis are presented. Number of replications are 1000, and n is the number of generated observations. Models (4.1) and (4.2)

Simulation study		*Test statistics*			
weak co-shifting	n	LM	LR	$Wald$	Rao
	200	0.052	0.053	0.06	0.045
	500	0.046	0.049	0.051	0.045
	1000	0.049	0.049	0.049	0.046
strong co-shifting					
	200	0.06	0.063	0.067	0.051
	500	0.055	0.055	0.056	0.051
	1000	0.047	0.047	0.048	0.046
weak co-seasonality					
	200	0.057	0.058	0.063	0.039
	500	0.067	0.068	0.069	0.061
	1000	0.054	0.055	0.055	0.054
strong co-seasonality					
	200	0.072	0.080	0.086	0.049
	500	0.053	0.056	0.056	0.045
	1000	0.053	0.053	0.055	0.052

Table A.2. Simulated power of the test for weak co-shifting. Rejection frequencies of the null hypothesis at the 5% level are presented in the table. Model (4.3). 1000 replications. n is the numbrer of observations.

Parameter variation	n	*Test statistics* LM	LR	$Wald$	Rao
$c_2 = 0.55$	200	0.21	0.22	0.23	0.19
	500	0.38	0.38	0.38	0.37
	1000	0.67	0.67	0.67	0.66
$c_2 = 0.58$	200	0.41	0.41	0.42	0.39
	500	0.79	0.79	0.79	0.78
	1000	0.97	0.97	0.97	0.97
$\gamma_2 = 15$	200	0.17	0.18	0.18	0.16
	500	0.32	0.32	0.32	0.30
	1000	0.55	0.55	0.55	0.55
$\gamma_2 = 20$	200	0.38	0.39	0.39	0.36
	500	0.72	0.73	0.73	0.71
	1000	0.95	0.95	0.96	0.95

Table A.3. Simulated power of the test for strong co-shifting. Rejection frequencies of the null hypothesis at the 5% level are presented in the table. Model (4.4).

$N = 1000$, n = number of generated observations

Parameter variation	n	*Test statistics* *LM*	*LR*	*Wald*	*Rao*
$c_2 = 0.55$	200	0.19	0.2	0.2	0.17
	500	0.34	0.34	0.34	0.32
	1000	0.59	0.59	0.59	0.59
$c_2 = 0.58$	200	0.39	0.4	0.42	0.36
	500	0.73	0.73	0.74	0.71
	1000	0.95	0.95	0.95	0.95
$\gamma_2 = 15$	200	0.18	0.18	0.18	0.16
	500	0.3	0.3	0.31	0.29
	1000	0.57	0.57	0.58	0.56
$\gamma_2 = 20$	200	0.39	0.4	0.41	0.36
	500	0.7	0.7	0.71	0.69
	1000	0.95	0.95	0.95	0.95
$a_2 = 1.2$	200	0.24	0.25	0.26	0.22
	500	0.48	0.49	0.49	0.41
	1000	0.8	0.8	0.8	0.8

Table A.4. Simulated power of the test for weak co-seasonality. Rejection frequencies of the null hypothesis at the 5% level are presented in the table. Model (4.5)

N = 1000, n = number of generated observations

Parameter variation	n	*Test statistics* LM	LR	$Wald$	Rao
$c_2 = 0.508$	200	0.11	0.11	0.12	0.081
	500	0.21	0.21	0.21	0.2
	1000	0.43	0.43	0.42	0.42
$c_2 = 0.512$	200	0.22	0.23	0.24	0.18
	500	0.48	0.48	0.48	0.46
	1000	0.77	0.77	0.78	0.77
$\gamma_2 = 10.7$	200	0.11	0.12	0.12	0.095
	500	0.18	0.18	0.18	0.16
	1000	0.36	0.36	0.36	0.36
$\gamma_2 = 11.1$	200	0.19	0.2	0.2	0.15
	500	0.41	0.41	0.41	0.38
	1000	0.71	0.71	0.71	0.70

Table A.5. Simulated power of the test for strong co-seasonality. Rejection frequencies of the null hypothesis at the 5% level are presented in the table. Model (4.6)

N = 1000, n = number of generated observations

Parameter variation	n	*Test statistics* LM	LR	$Wald$	Rao
$c_2 = 0.65$	200	0.12	0.13	0.15	0.11
	500	0.18	0.19	0.2	0.17
	1000	0.35	0.36	0.37	0.35
$c_2 = 0.8$	200	0.28	0.26	0.28	0.20
	500	0.44	0.44	0.45	0.42
	1000	0.69	0.70	0.70	0.68
$\gamma_2 = 100$	200	0.11	0.13	0.14	0.086
	500	0.3	0.32	0.33	0.28
	1000	0.63	0.64	0.65	0.62
$\gamma_2 = 1000$	200	0.15	0.16	0.18	0.12
	500	0.4	0.41	0.43	0.37
	1000	0.79	0.79	0.8	0.78
$\lambda_{202} = 3.5$	200	0.4	0.42	0.44	0.36
	500	0.55	0.55	0.56	0.53
	1000	0.8	0.81	0.81	0.79
$\lambda_{212} = 1.6$	200	0.21	0.23	0.25	0.17
	500	0.41	0.43	0.43	0.39
	1000	0.77	0.77	0.77	0.76
$\lambda_{222} = 1.5$	200	0.22	0.24	0.26	0.18
	500	0.48	0.49	0.51	0.46
	1000	0.83	0.83	0.84	0.83
$\lambda_{232} = -1.5$	200	0.25	0.27	0.29	0.2
	500	0.62	0.62	0.64	0.59
	1000	0.93	0.93	0.93	0.92

Table A.6. Test for shifting of four-quarter difference and changing seasonality of first difference of Japanese income and consumption. The test is applied by constructing AR(p) models and tests the constancy of the intercept in the different series

Tested data series	LM_1	LM_2	LM_3
4th difference of Japanese income, $AR(13)$	0.0004	0.029	0.0059
4th difference of japanese consumption, $AR(13)$	0.0067	0.07	0.12
1st difference of japanese income, $AR(10)$	$2.7e^{-7}$	0.004	0.001
1st difference of japanese consumption, $AR(10)$	0.032	0.13	0.1

The tests are based on Taylor expansion as in Lin and Teräsvirta (1994)

Table A.7. Test for co-shifting of four-quarter difference and co-seasonality of first difference of Japanese income and consumption.

	Test statistics (p-values in parentheses) $Wald$
weak co-shifting	4.49(0.11)
strong co-shifting	3.71(0.29)
weak co-seasonality	9.69(0.0079)
weak co-seasonality	43.94(0.00)

Table A.8 Estimated model for co-shifting under H_1
four-quarter difference of Japanese consumption $= \Delta_4 y_1$
four-quarter difference of Japanese income $= \Delta_4 y_2$

$R^2 = 0.83 \quad s = 0.013$

$$\Delta_4 y_{1,t} = \underset{0.013}{0.061} + \underset{0.087}{0.49}\Delta_4 y_{1,t-1} + \underset{0.092}{0.27}\Delta_4 y_{1,t-2} - \underset{0.085}{0.60}\Delta_4 y_{1,t-4} + \underset{0.097}{0.28}\Delta_4 y_{1,t-6} - \underset{0.089}{0.17}\Delta_4 y_{1,t-8}$$
$$-\underset{0.0072}{0.035}\left[1 + \exp\left\{-\underset{14.98}{12.34}\left(t^* - \underset{0.0077}{0.40}\right)/0.084\right\}\right]^{-1} + \widehat{\varepsilon}_{1t}$$

$R^2 = 0.63 \quad s = 0.024.$

$$\Delta_4 y_{2,t} = \underset{0.013}{0.10} + \underset{0.098}{0.14}\Delta_4 y_{2,t-3} - \underset{0.10}{0.29}\Delta_4 y_{2,t-4}$$
$$-\underset{0.012}{0.070}\left[1 + \exp\left\{-\underset{3.96}{4.54}\left(t^* - \underset{0.021}{0.44}\right)/0.084\right\}\right]^{-1} + \widehat{\varepsilon}_{2t}$$

Log Likelihood $= 504.86$

Table A.9 Estimated model for co-seasonality under H_1
first difference of Japanese consumption $= \Delta y_1$
first difference of Japanese income $= \Delta y_2$

$R^2 = 0.98 \quad s = 0.013$

$$\Delta yc_{1,t} = \underset{0.015}{0.18} - \underset{0.15}{0.16}\Delta yc_{1,t-1} - \underset{0.11}{0.12}\Delta yc_{1,t-5} - \underset{0.016}{0.13}d_1 - \underset{0.039}{0.17}d_2 - \underset{0.04}{0.3}d_3$$
$$\left(-\underset{0.02}{0.07} - \underset{0.023}{0.07}d_1 + \underset{0.018}{0.045}d_2 - \underset{0.032}{0.091}d_3\right)\left[1 + \exp\left\{-\underset{0.40}{0.76}\left(t^* - \underset{0.061}{0.32}\right)/\widehat{\sigma}_{y_1}\right\}\right]^{-1} + \widehat{\varepsilon}_{1t}$$

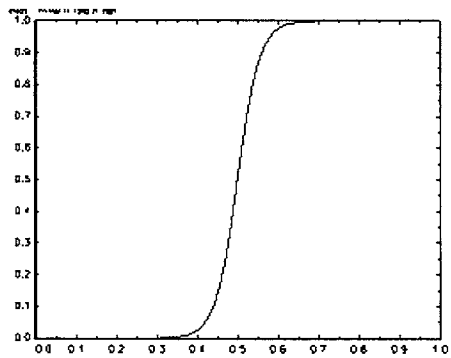
$R^2 = 0.99 \quad s = 0.021$

$$\Delta yi_{2,t} = \underset{0.023}{0.35} - \underset{0.1}{0.55}\Delta yi_{2,t-1} - \underset{0.12}{0.33}\Delta yi_{2,t-2} - \underset{0.059}{0.32}d_1 - \underset{0.044}{0.36}d_2 - \underset{0.019}{0.58}d_3 +$$
$$\left(-\underset{0.024}{0.075} - \underset{0.032}{0.01}d_1 + \underset{0.03}{0.17}d_2 - \underset{0.03}{0.02}d_3\right)\left[1 + \exp\left\{-\underset{0.34}{94}\left(t^* - \underset{0.032}{0.54}\right)/\widehat{\sigma}_{y_2}\right\}\right]^{-1} + \widehat{\varepsilon}_{2t}$$

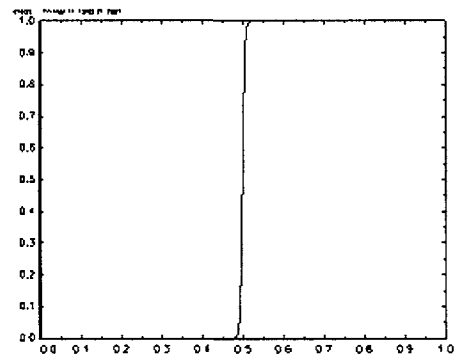
Log Likelihood $= 531.02$

B. Figures

Figure B.1. Transition function as a function of the transition variable t of an LSTAR(1) model, for different speed of transition.

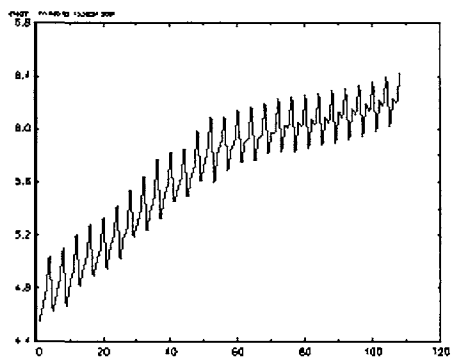


$\gamma = 10$

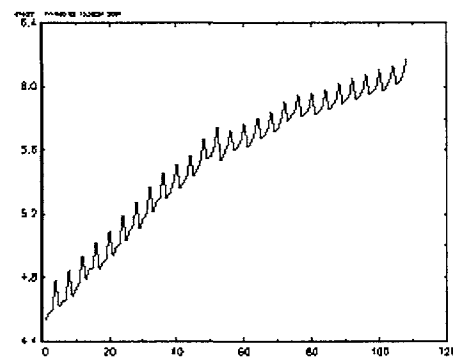


$\gamma = 100$

Figure B.2. Japanese income and Japanese consumption.

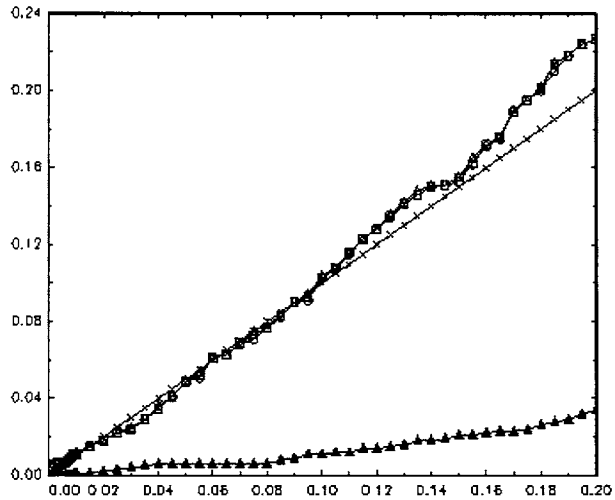


Japanese income

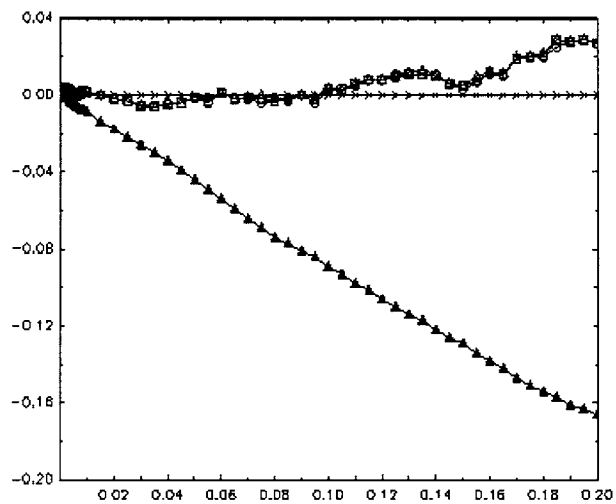


Japanese consumption

Figure B.3. P-value plots, weak co-shifting.
LM-test (circled), LR-test (quadratic), Wald-test (triangled),
Rao-test (filled triangled) and 45°/zero-line (crossed).
1000 observations.

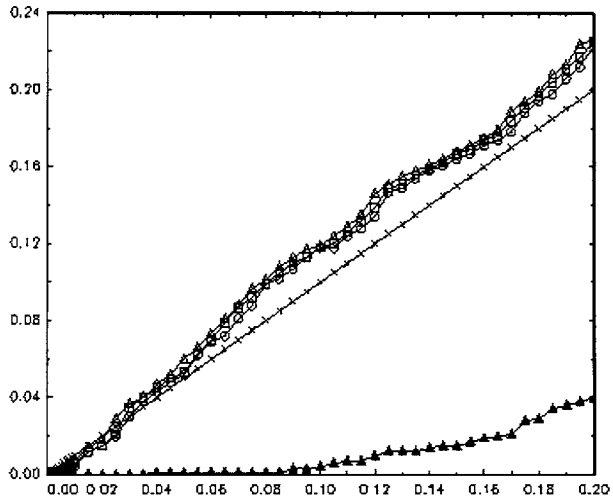


Truncated P value plot

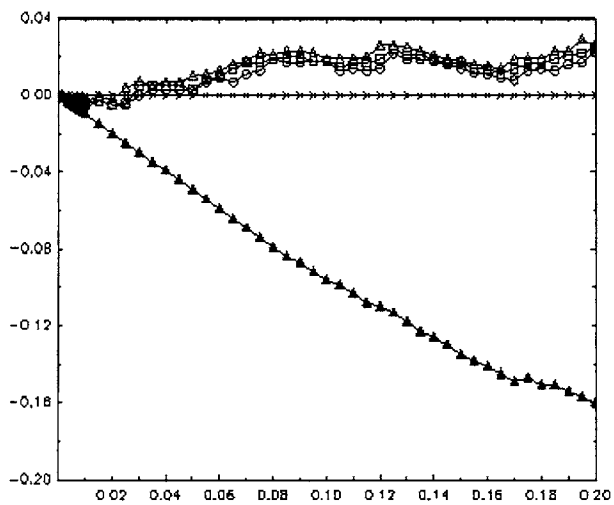


Truncated P value discrepancy plot

Figure B.4. P-value plots, weak co-shifting.
LM-test (circled), LR-test (quadratic), Wald-test (triangled)
Rao-test (filled triangled) and 45°/zero-line (crossed).
200 observations.



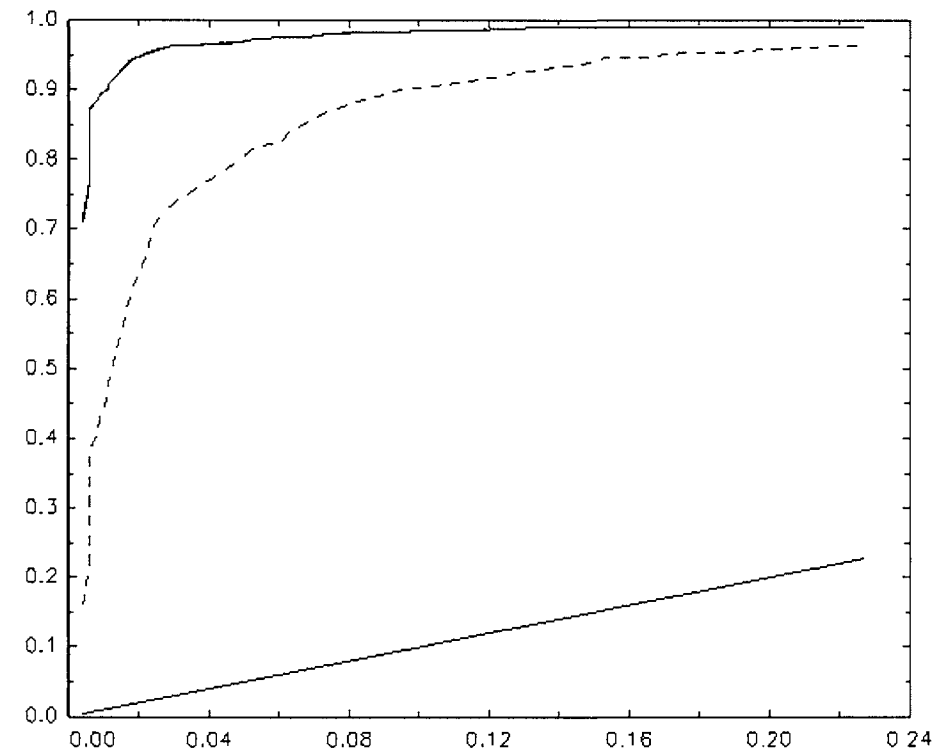
Truncated P value plot



Truncated P value discrepancy plot

Figure B.5. Size-power curves, weak co-shifting.
The size power curves for the *LM-, LR-, Wald-test*
lies so close that they cannot be distinguished from each other. *The dashed line is the Rao-test.*

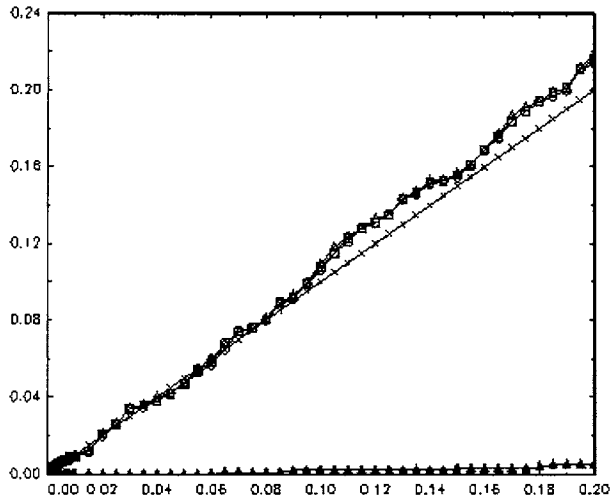
1000 observations.



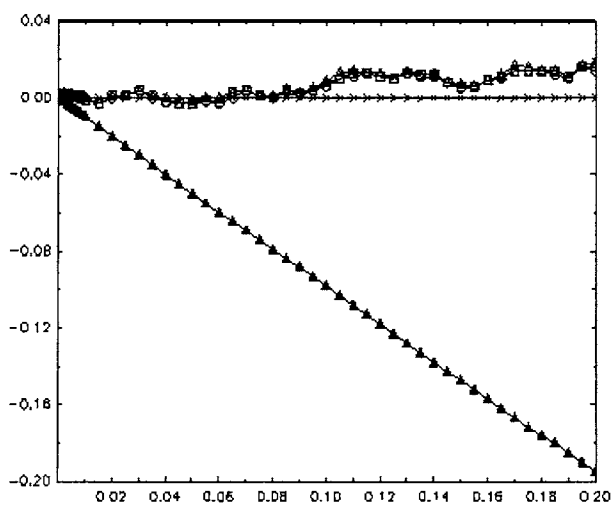
$c_2 = 0.58$

Figure B.7. P-value plots, strong co-shifting.
LM-test (circled), LR-test (quadratic), Wald-test (triangled) and 45°/zero-line (crossed).

1000 observations.



Truncated P value plot

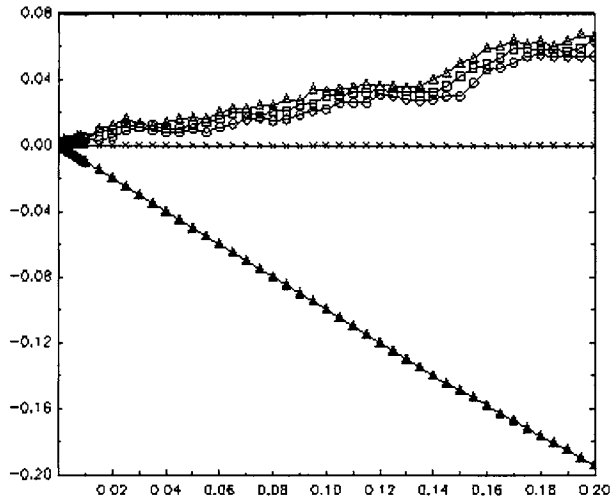


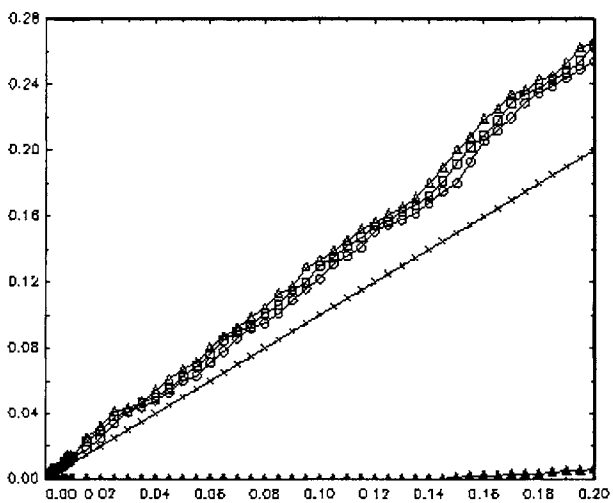
Truncated P value discrepancy plot

Figure B.8. P-value plots, strong co-shifting.
LM-test (circled), LR-test (quadratic), Wald-test (triangled) and 45°/zero-line (crossed).

200 observations.

Truncated P value plot

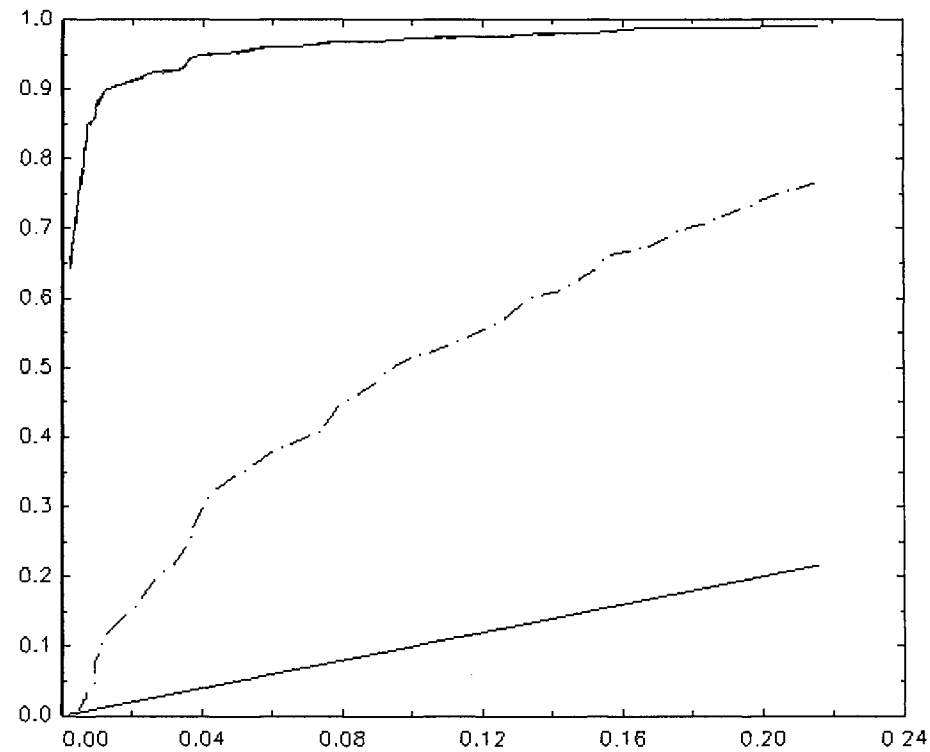




Truncated P value discrepancy plot

Figure B.9. Size-power curves, strong co-shifting.
The size power curves for the *LM-, LR-, Wald-test*
lies so close that they cannot be distinguished from each other. *The dashed line is the Rao-test.*

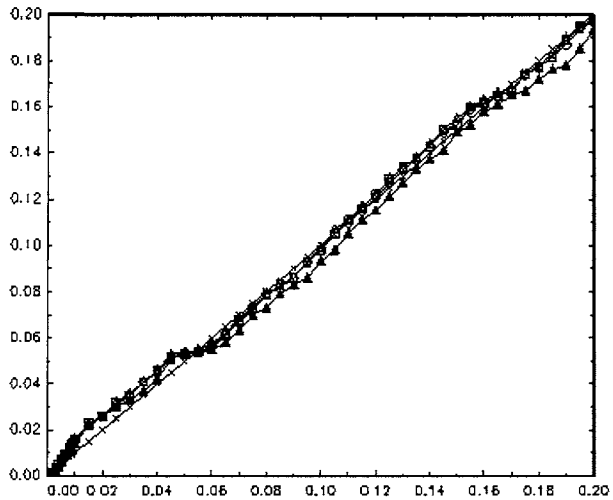
1000 observations.



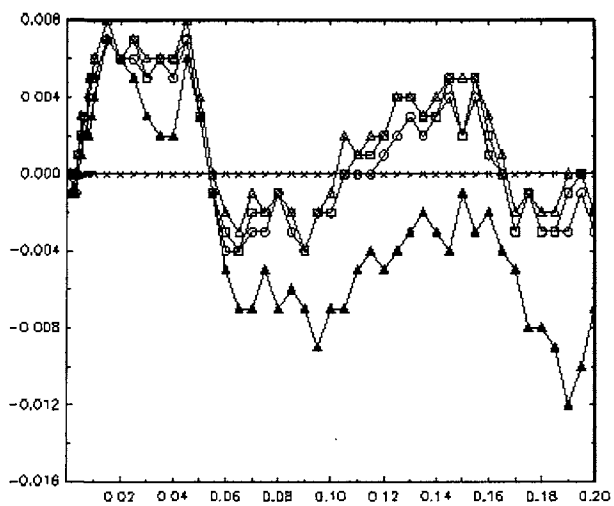
$c_2 = 0.58$

Figure B.11. P-value plots, weak co-seasonality.
LM-test (circled), LR-test (quadratic), Wald-test (triangled) and 45°/zero-line (crossed).

1000 observations.



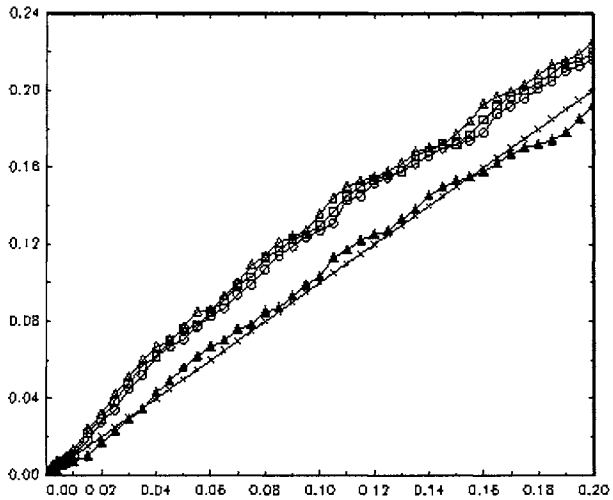
Truncated P value plot



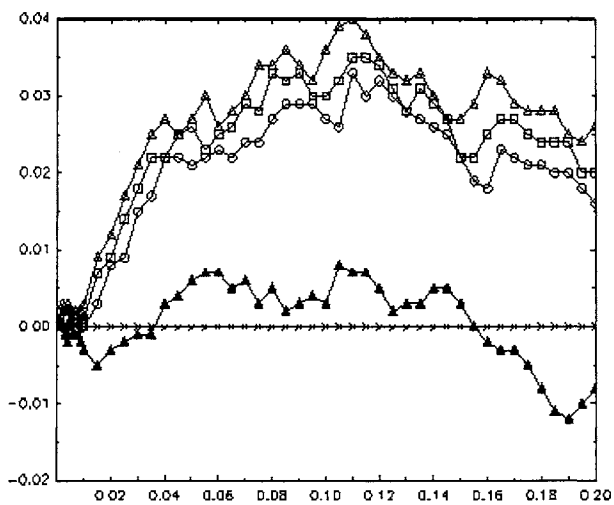
Truncated P value discrepancy plot

Figure B.12. P-value plots, weak co-seasonality.
LM-test (circled), LR-test (quadratic), Wald-test (triangled) and 45°/zero-line (crossed).

200 observations.



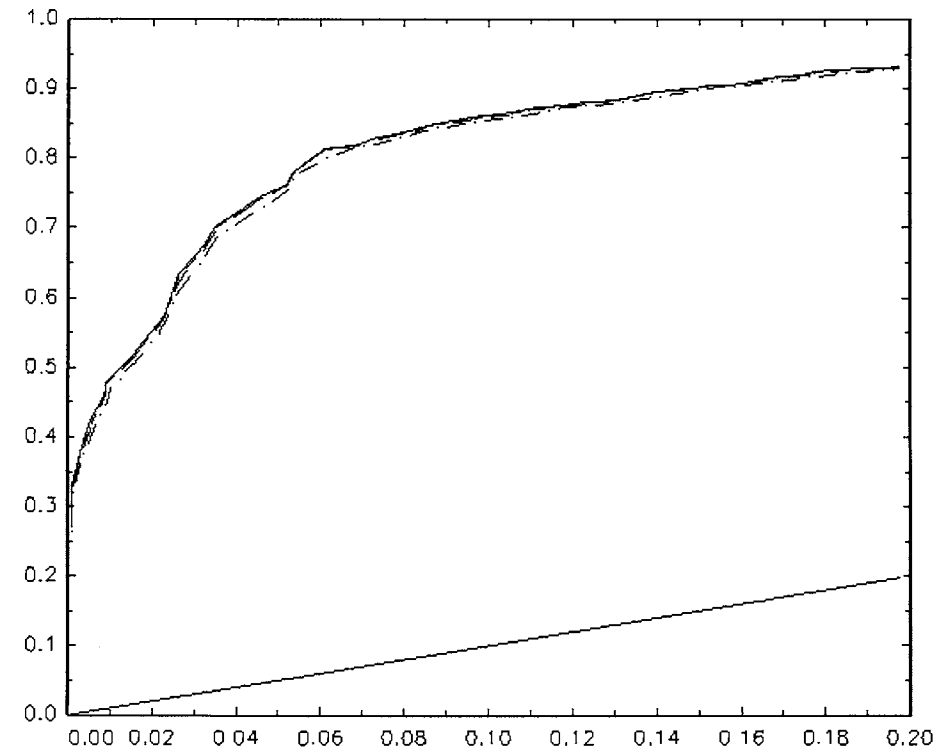
Truncated P value plot



Truncated P value discrepancy plot

Figure B.13. Size-power curves, weak co-seasonality.
The size power curves for the tests *LM-test, LR-test, Wald-test*
lies so close that they cannot be distinguished from each other. *The dashed line is the Rao-test.*

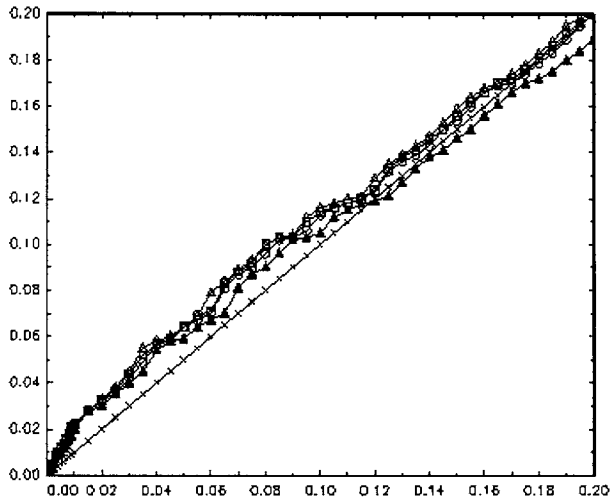
1000 observations.



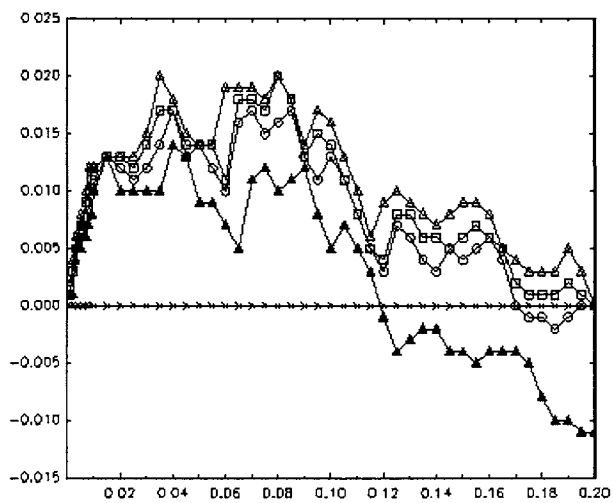
$c_2 = 0.512$

Figure B.15. P-value plots, strong co-seasonality.
*LM-test (circled), LR-test (quadratic), Wald-test (triangled) and 45°/*zero-line (crossed).

1000 observations.



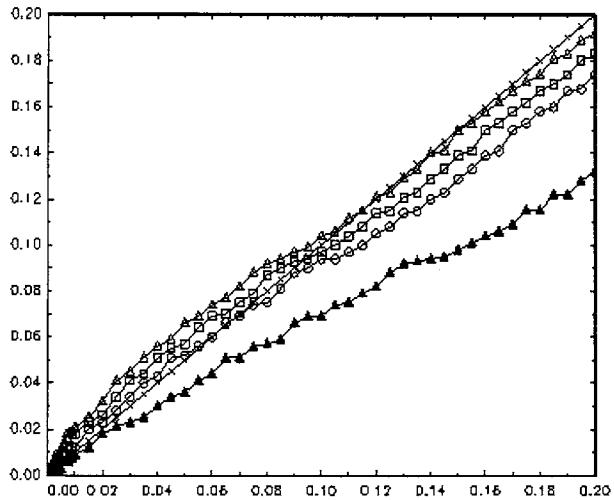
Truncated P value plot



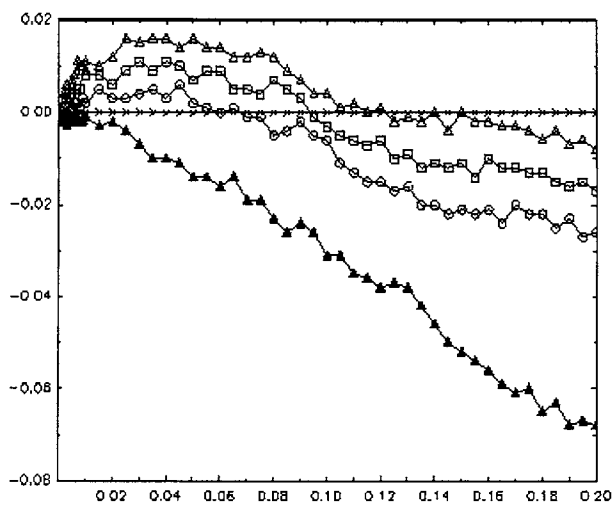
Truncated P value discrepancy plot

Figure B.16. P-value plots, strong co-seasonality.
*LM-test (circled), LR-test (quadratic), Wald-test (triangled) and 45°/*zero-line (crossed).

200 observations.



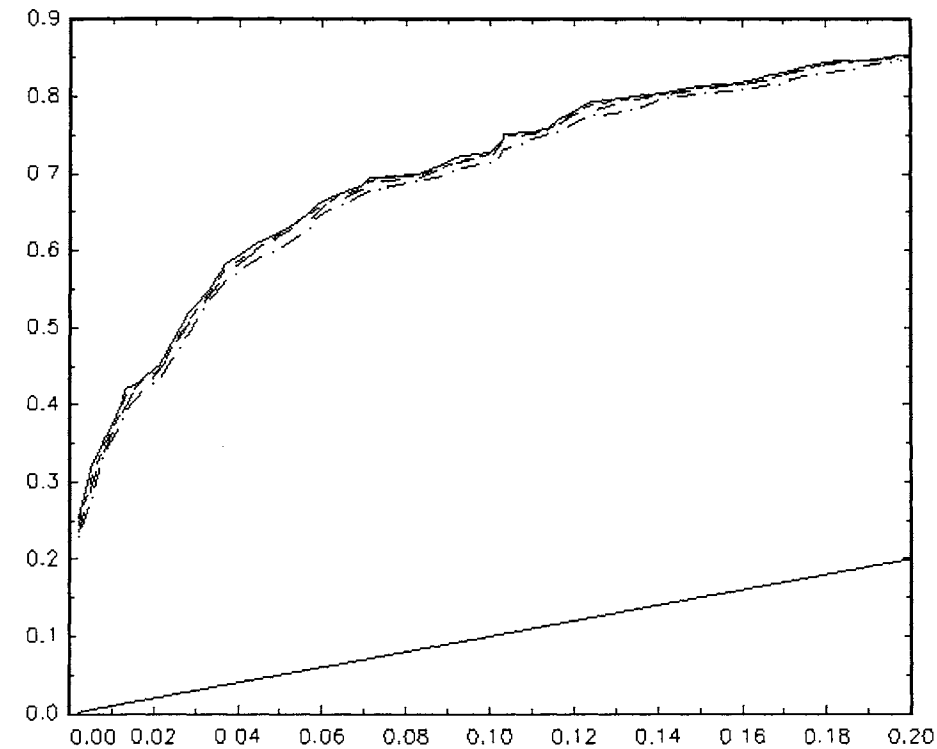
Truncated P value plot



Truncated P value discrepancy plot

Figure B.17. Size-power curves, strong co-seasonality.
The size power curves for the tests *LM-test, LR-test, Wald-test*
lies so close that they cannot be distinguished from each other. *The dashed line is the Rao-test.*

1000 observations.



$c_2 = 0.8$

Measuring asthma patient satisfaction
in Sweden using Partial Least Squares

Anna Ovanfors
Department of Economic Statistics, Stockholm School of Economics
Box 6501, SE-113 83 Stockholm, Sweden
October 14, 2004

Abstract

This paper examines the relationships between different aspects involved in asthma treatment. Each aspect's impact on overall patient satisfaction with their asthma treatment (Patient Satisfaction Index, PSI) is analysed. We also study how outcome variables such as "compliance with physician's recommendations", "Health-related quality of life" and "resource use" are affected by the degree of patient satisfaction. The results from this study refer to the asthma patients as a group but not necessarily to each patient as an individual.

The statistical technique applied for this analysis is Partial Least Squares (PLS), which is well suited for structural equation modelling when the focus is on identifying the most important characteristics involved in the asthma treatment, aimed at making powerful recommendations for optimised improvements. The suggested generic model is tested on 599 respondents from a questionnaire survey. The structure of the suggested model is well supported by the data.

Key words. asthma, compliance, self-management, health-related quality of life, Patient Satisfaction Index (PSI), Partial Least Squares (PLS)

1 Introduction

Asthma and asthma treatment are evaluated from a number of different perspectives, including clinical effects, health-related quality of life (HRQL hereafter), costs and cost-effectiveness. The medical and health economic studies are often designed as randomised clinical experiments (see for example Anderson et al., 2000; Zetterström et al., 2000; Berggren and Ekström, 1999). These types of studies show clinical effects and cost-effectiveness of drug treatment under well-controlled conditions.

A distinguishing characteristic in asthma treatment is the importance of "Compliance with the physician's recommendations", and "Self-management" of the overall asthma treatment. Gibson et al. (2000) show in a meta-analysis that asthma treatment which focuses on "Self-management" number of days spent by the patients at hospital and the number of days they are absent from work, compared to standard asthma treatment (see also Liljas and Lahdensuo, 1997). ""

A number of Swedish studies show that "asthma health care centres" reduce the number of emergency visits, asthma attacks, days in hospital, antibiotic treatment and days of absence from work (see Lorentzson, 1995; Dahlberg et al., 1997; Lisspers and Österlund Efraimsson, 1998). One problem, though, with these studies is that they are designed as before/after studies, and that the number of patients is relatively small, which leads to difficulties when drawing conclusions that are definite (see Swedish Council on Technology Assessment in Health Care, 2000; National Board of Health and Welfare, 2003).

The compilation by the Swedish Council on Technology Assessment in Health Care (2000) and the National Board of Health and Welfare (2003) regarding knowledge of treatment of asthma and Chronic Obstructive Pulmonary Disease (COPD hereafter) also includes drug treatment evaluations. Thus, there are many factors playing an important role in the results of the asthma treatment of adults as well as regarding its effectiveness and cost. However, the documentation and the knowledge are insufficient. The Swedish Council on Technology Assessment in Health Care (2000) and the National Board of Health and Welfare (2003) conclude that there is a need for surveying the best way of organizing the treatment of patients with asthma and COPD. There is also a need for more knowledge about which components involved in the asthma treatment (e.g. the physician, asthma nurse, education about asthma as a disease, availability) are of greatest importance for an optimal treatment of the patient.

The main purpose of this study is to identify the most important factors involved in the treatment of asthma, and how asthma patients rate these factors. For this purpose, we analyse patients' perceptions of the separate factors involved in asthma treatment and the impact of each factor on the overall satisfaction with their asthma treatment: the Patient Satisfaction Index (PSI hereafter). These results will then be linked to "Compliance with the physician's recommendations" about the asthma treatment, degree of "Self-management" in the asthma treatment, resulting HRQL, and "Sesource use". A secondary objective of the study is to investigate asthma treatment at "asthma health care centres" compared to asthma treatment at ordinary health care centres. An "asthma health care centre" was defined as a centre where patients can make appointments and call regularly; which can perform spirometry and has a structured way of working based on a programme of care; does PEF measurement, reversibility testing, planning of care, and gives information on inhalation technique.

The plan of the paper is as follows. Firstly, we will discuss the data collection and the design of the study. Secondly, we give a short summary of the PLS method chosen for this study. Thirdly, the empirical results of the PLS analysis will be discussed. In section 4, the empirical results will be summarized. In the fifth section, we will shortly discuss the patients' resource use, and finally we will conclude the paper.

2 Data collection

2.1 Comprehensive design of the study

Based on qualitative interviews[1], a questionnaire was designed that was tested on two focus groups for verification (see appendix Table A1 for the questionnaire). The next phase was a multi-centre study based on the questionnaire, where the patients were told to rate the activities defining their asthma treatment. Each question was evaluated on a 1 to 10 scale by the patients. In these types of surveys the questions that are graded like this and the constructed factors are interpreted as interval scales (i.e. equidistance is assumed between the grades in the scale (see Fornell, 1992; Fornell and Cha, 1994; Fornell et al. 1996; EPSI Technical Committee, 1998; Anderson and Fornell, 2000).

Based on the interviews and the questionnaire, a hypothetical model was constructed. In this model, the questions that are closely related, excluding the background questions, are grouped into factors, or so-called latent variables (see appendix Table A3 for the first two eigenvalues for each factor, respectively, where the unidimensionality is verified by factor analysis). If the first eigenvalue is above one and the second eigenvalue is less than one (no more than 0.8 preferably), the factor is considered to be unidimensional.

Among the factors are the Patient Satisfaction Index (PSI), which is a measure based on three questions that are weighted together as an index (see section 3.2): "Overall how satisfied are you with your asthma treatment", "How well does your asthma treatment meet your expectations" and "Imagine an asthma treatment that is perfect in every sense. How close is your current asthma treatment to this ideal".

PSI is a measure of how satisfied the patients are with their overall asthma treatment. In Figure 1 the empirical model is described where other factors involved in the asthma treatment are affecting PSI: "The pysician's manner", "The nurse's manner", "Availability of the health centre" and "The drug". The purpose is to identify the most important factor involved in the overall asthma treatment in order to make concrete recommendations for optimised improvements in PSI. This will in turn lead to more compliance with physician's recommendations, better HRQL and less resource use for the patients. Therefore, the patients' perceptions of the separate factors involved in the asthma treatment and the impact of each factor on PSI are analysed.

2.2 Study population

Patients were recruited from 17 health care centres: 11 in Region Skåne and 6 in Stockholm County, respectively. Questionnaires were sent out to 796 patients, in total, of which 644 patients responded. Of these patients 20 were of the wrong age according to the inclusion criteria below, and were therefore excluded. Further, 26 patients who only answered less than

[1] Interviews were held with physicians, politicians and civil servants from Region Skåne and from Stockholm County Council, the Asthma- and Allergy Association, experts at AstraZeneca, and 10 patients.

two-thirds of the questions in the questionnaire were also excluded. The final result was 599 analysed respondents.

The following patients are part of the study population:

- Patients with an asthma diagnosis, treated at health care centres in Region Skåne, and in Stockholm County.
- The patients should have been under the given prescription for at least 6 months:
  Group A = glucocorticoid + short-term-acting beta-2-agonists in one inhaler.
  Group B = glucocorticoid + long-term-acting beta-2-agonists in separate inhalers.
  Group C = glucocorticoid + long-term-acting beta-2-agonists in one inhaler.
  Note that patients within group B and C probably have a more severe type of asthma as compared to A.
- Patients who can read and write in Swedish without any problems.
- Patients between 18 and 65 years of age.

Patients with a COPD diagnosis, patients with a known abuse problem and patients taking part in any other scientific study at the same time were, however, excluded from the study population.

The study was approved by the Research Ethics Committees at Lund University and Karolinska Institutet, Stockholm.

3 Methodology

3.1 Choice of analysis technique

In order to identify the most important factors involved in the asthma treatment and how asthma patients rate these factors, a generic structural equation model was constructed. The empirical model can be seen in Figure 1. Different factors involved in asthma treatment are on the left-hand side of the model; so-called control variables of PSI. On the right-hand side of the model we have goal variables – "Compliance with physicians recommendations", "Asthma control" and patients HRQL. The hypothesis is that increased PSI will lead to increased "Compliance with physicians recommendations", more "Asthma control" and higher HRQL, that in turn will have positive effects for the costs to society.

Wold (1966) proposed an analysis technique, Partial Least Squares (PLS), which has evolved in many different directions since then. The CFI Group carried out the statistical analysis in this study, using the methodology developed by Claes Fornell (see Fornell and Cha, 1994). This method is prepared for estimating the causal relations between respondents' perceptions of a) different aspects of their situation, b) their general situation and c) the behavioural outcome. The method has been carefully evaluated and finally chosen as the method of measuring perceived quality by customers at branch level in Europe (EPSI) (see Fornell, 1992; EPSI Technical Committee, 1998), and in USA (ACSI) (Fornell et al. 1996; Anderson and Fornell, 2000; Johnson et al. 2001).

The PLS-technique, which is not based upon any distributional assumptions in the variables, is well suited for structural equation modelling when the focus is on predictive power in the model (see for example Jöreskog and Wold, 1982). This is an important factor in favour of the PLS-technique as apposed other techniques, because in practice the distribution of variables is often unknown, or greatly differs from a normality distribution. In our data this is true for all of the variables, which are skewed towards the higher scores and far from normally distributed. In those cases, analysis by covariance structure modelling, as in LISREL, is not feasible. Thus, several recent studies use the PLS-methodology (Westlund, 1994; Eckerlund et al. 1997; Henriksson et al. 1997; Westlund, et al. 1998; Eckerlund et al. 2000).

The PLS-method, that is a variance-based analysis technique, is also different from covariance-structure-modelling, such as LISREL, in the way that the PLS-method delivers scores on factor level, so-called latent variables, while covariance-structure-modelling does not. The factors are estimated by means of weighted relations (see, for example, Fornell and Cha, 1994). The causal relationships between the factors in the model are optimised according to the iterative PLS estimation that is summarized below (see also Dijkstra, 1983). Questions that do not contribute to the analysis are left out (see section 4.2.1 Reliability and Construct validity for explanation).

3.2 Model structure

Each factor in the hypothetical model; for example, "Competence of the physician" (see Table A1 in the appendix) is a so-called latent variable. Latent variables are not directly observable but estimated through weighted relations by a group of statements, or manifest variables, from the questionnaire. The inner relations between the factors are assumed to be linear, as represented by arrows in the model shown in Figure 1.

PLS modelling consists of three parts, (see Fornell and Cha, 1994): inner relations, outer relations and weighted relations. The inner relations, or the relations between the factors, can be written as follows:

$$\eta = \mathbf{B}\eta + \mathbf{\Gamma}\xi + \zeta \tag{1}$$

where vectors of endogenous and exogenous factors are η and ξ respectively, while ζ is a residual vector. $\mathbf{B}$ and $\mathbf{\Gamma}$ are matrices of regression coefficients.
The outer relations describe the relations between the factors and the directly observed manifests, or statements, which form the factors:

$$\begin{aligned} y &= \mathbf{\Lambda}_y\eta + \varepsilon_y \\ x &= \mathbf{\Lambda}_x\xi + \varepsilon_x \end{aligned} \tag{2}$$

where η and ξ are the vectors of endogenous and exogenous factors respectively, y and x are the observed indicators, or statements, the so-called manifest variables of η and ξ respectively. $\mathbf{\Lambda}_y$ and $\mathbf{\Lambda}_x$ are matrices of loadings, which relate the factors to their measures. Finally, ε_x and ε_y are the residuals, here assumed white noise.

The outer relations are also ruled by predictor specification:

$$E[y|\eta] = \Lambda_y \eta$$
$$E[x|\xi] = \Lambda_x \xi \tag{3}$$

where $E[y|\eta]$ and $E[x|\xi]$ are the means, or the expected values of the observed indicators y and x respectively, given the vectors of endogenous and exogenous factors η and ξ respectively.

The last relation required in the PLS-estimation is the weight relation. In PLS, each case value of the factors is estimated as follows:

$$\hat{\eta} = \omega_\eta y$$
$$\hat{\xi} = \omega_\xi x \tag{4}$$

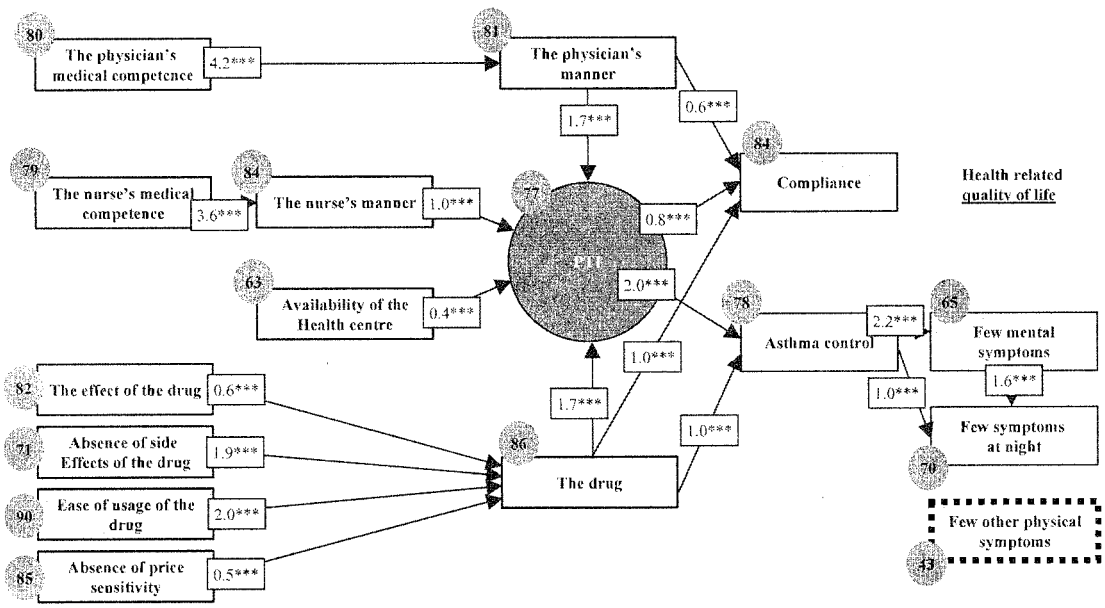
where ω_η and ω_ξ are the weights.

In PLS, some parameters are estimated by means of multiple regression technique while other parameters are held fixed at the time. ζ and ε, are handled separately (see Wold, 1966; Cassel et al., 1999) and thus approximations of the estimates are delivered successively, part by part, in an iterative process. Estimation, thus, aims at minimizing $Var[\zeta]$ for each endogenous factors and minimizing $Var[\varepsilon]$ for each outer relation.

4 Empirical results

The operational model is shown in Figure 1. The model should be interpreted in the following way: "The physician's manner" has, for example, a regression coefficient of 1,7 on PSI; this means that the impact of a score increase in "The physician's manner" by 5 units[2] will result in a score increase of the PSI by 1,7 units, ceteris paribus. Hence, 5 units score increase in "The physician's medical competence" will result in 4.2 units score increase in "The physician's manner".

Figure 1. Empirical model of the Patient Satisfaction Index
Scores are presented in circles on the left hand side of the factors, and OLS regression coefficients in the boxes to the right (where **, *** means significant at 5%- and 1%-level respectively).



Note that a higher score means fewer problems. See questions in appendix Table 1.
Estimated coefficients by ordinary PLS- and double PLS-technique are presented in Appendix Table A5.

[2] The choice of describing the impact of a five-unit score increase instead of a one-unit score increase is because a five-unit increase is a reasonable goal, whereas a one-unit score increase would not be significantly notable.

4.1 Implications from the empirical model

Looking at Figure 1 we can see that to attain an increased PSI, score increases in "The physician's manner" and "The drug" would make the greatest contribution, since both factors have the largest impact (regression coefficients) of 1.7 on PSI. "The nurse's manner" also has a relatively greatest impact, whereas the impact of the "Availability of the health centre", although significant, is relatively small.

The "The physician's medical competence" and the "The nurse's medical competence" impact the "The physician's manner" and the "The nurse's manner" respectively. This means that the "The physician's medical competence" and the "The nurse's medical competence" have only an indirect impact on PSI.

"Ease of usage of the drug" and "Absence of side effects" have the greatest impact on "The drug", whereas "The effect of the drug " and "Absence of price sensitivity" have relatively little impact.

"The drug" is the factor having the greatest impact on "Compliance". As "Ease of usage of the drug" and "Absence of side effects" to large extent determine whether the patients are satisfied with "The drug" they indirectly have a great impact on "Compliance". "The physician's manner" also has a great impact on "Compliance".

"Asthma control" is determined by PSI and "The drug". Furthermore, "Asthma control" has a considerable impact on HRQL factors. Good "Asthma control" leads to "Few mental symptoms" and "Fewer symptoms at night". "The drug" and the "The physician's manner" also have an indirect effect on compliance through the PSI.

It is interesting to notice that PSI is relatively high while at the same time the patients rate their HRQL relatively low. Among the HRQL factors, "Few other physical symptoms" is the factor having the lowest score of 43.

Ranking the different factors according to scores and regression coefficients/impacts can summarize the results and implications of the empirical model. Factors with a relatively low score but relatively large impact are set to "priority in first hand", which would be "The physician's manner" and "The physician's medical competence". Thus, even if the scores are not low for either "The physician's manner" or "The physician's medical competence" they are relatively lower than the scores for other factors. This means that a further score increase in "The physician's manner" and "The physician's medical competence" would have a relatively high impact on PSI. Furthermore, we can conclude that it is important to maintain the patient's perception of "The drug" and "The nurse's manner". These factors have higher scores compared to other factors in the model, but improvements will still have a considerable effect on PSI.

This does not imply, however, that all other factors are unimportant. Firstly, all factors included in the empirical model (Figure 1) are important, or else we would not have taken them into account. Secondly, the model is dynamic over time, which means that this is the case only for the point in time when the survey was conducted. Once you start to work with the improvements in asthma treatment, and the patients notice those improvements, the impact structure starts to change. Likewise, if you forget to work with improvements in some factors that have been given low priority at one point in time, to the degree that the patients

may think these factors are below a certain 'hygienic level', then these factors may be high priorities next time you measure. The implication is that you need to follow up these types of analyses continuously in order to make an effective action plan over time.

Patients treated at asthma health care centres compared to patients treated at ordinary health care centres, and patients in the different treatment groups A, B and C, experience their asthma treatment differently. The results show that patients treated at asthma health care centres are significantly more satisfied with the overall asthma treatment; that they are significantly more "Compliant" and have significantly better "Asthma control" compared to patients treated at an ordinary health care centre, as can be seen from Table 1.

Table 1. Differences in scores for patients treated at asthma health care centres versus patients treated at ordinary health care centres.
p-values, at 0,05 or below, on t-tests vs. the other group are presented in parenthesis.

Factors	Asthma health care centres	Ordinary health care centres
PSI	78 (0,05)	74
Compliance	85 (0,000)	81
Asthma control	80 (0,000)	76
Few mental symptoms	66	64
Few symptoms at night	70	71
Few other physical symptoms	44	-

Furthermore, from Table 2 we see that patients in treatment group C are significantly more "Compliant" than patients in treatment group A, and patients in treatment group B rate their HRQL (i.e. "Few mental symptoms", "Few symptoms at night" and "Few other physical symptoms") lowest, while treatment groups A and C rate their HRQL rather equally. However, patients within group B and C probably have a more severe asthma as compared to group A.

Table 2. Differences in scores between treatment groups
p-values, at 0,05 or below, on t-tests vs. the other groups are presented in parenthesis. p_a and p_b means significantly different vs. group A and B respectively.

Factors	**Treatment groups**		
	A	**B**	**C**
PSI	76	78	77
Compliance	82	84	85(p_a 0,05)
Asthma control	80	78	78
Few mental symptoms	66 (p_b 0,05)	63	67(p_b 0,05)
Few symptoms at night	69	68	72(p_b 0,05)
Few other physical symptoms	45 (p_b 0,05)	41	44(p_b 0,05)

4.2 Tests of the predictive power of the empirical model

Predicted, oriented tests that are well suited for evaluation of a model estimated by PLS-technique are communality, structural prediction, validity, redundancy and operational variance (see Lohmöller 1989). These tests are non-parametric; tests that are not based on any distributional assumptions in the variables, since PLS does not rely on statistical distribution (see Wold, 1982).

Predictions are made in each of the three PLS-estimation steps described above; thus, we need several different measures to evaluate overall predictive power in the empirical model. Below we give a summary of the measures we have used.

4.2.1 Reliability and construct validity

The manifests can be predicted from their respective factors. Cronbach's alpha provides an overall measure of the reliability, i.e. the degree to which the measures are free from errors, and therefore yield consistent results (Fornell and Cha, 1994). A Cronbach's alpha of around 0.6 and above is considered satisfactory. In our study (see Table A3 in appendix), Cronbach's alpha is above 0.75 for all factors, except for "The effect of the drug" which has a Cronbach's alpha of 0.59.

Average Variance Extracted (AVE) is the part of the variation that has been captured by the corresponding factor in proportion to the amount of variation due to measurement error, (see Fornell and Cha, 1994). AVE above 0.5 is considered adequate. The AVE in our study (see Table A3 in appendix) is above 0.7 for all factors, except for the "Few mental symptoms" factor that has an AVE of 0.6.

4.2.2 R^2: the amount of total variance captured by the model

In regular regression analysis, R^2 is the total amount of variance in the endogenous variable that has been explained by the model. Each inner structural part in the empirical model has been estimated by OLS, and each inner relation is given a measure of R^2 (see Fornell and Cha, 1994). In Appendix, Table A2, R^2 is presented for each estimated inner relation. For the first four relations the R^2 is relatively high, while it is lower for the four following relations (on the right hand side of the empirical model in Figure 1), but still high enough for being a statistical relation.

However, it should be noted that R^2 is not a measure of the overall quality in a structural relation. If, for example, the endogenous variable contains a relatively small amount of variation, the R^2 will also be relatively small due to the fact that a small amount of variation is often rather difficult to 'capture' (see Gujarati, 1995; Ramanathan, 1998).

4.2.3 Measures of the degree of multicollinearity in the empirical model

The above measures must be completed by an assessment of the degree of multicollinearity in the model. The condition index is a measure of the degree of multicollinearity in the structural relation (see Gujarati, 1995). In our study (see Table A4) the largest condition index is less than 21 for each structural relation. If the largest condition index in a structural relation is greater than 30, the degree of multicollinearity can be said to be strong, and not desirable.

Furthermore, we can see in Figure 1 above that all regression coefficients, or impacts, are significant at the 1% or the 5% level, which also indicates that the multicollinearity is not severe.

4.2.4 Further multicollinearity reduction

To reduce the multicollinearity further, and to increase the precision in the estimated regression coefficients, it is possible to perform a PLS-estimation that extracts the information that is relevant from the factors in the empirical model. The information extracted creates new variables, from which direct regression coefficients, or impacts, are estimated instead. The impacts from the factors to the endogenous variable can easily be calculated through the indirect impacts from the factors to the new variables by multiplication. This method combined with ordinary PLS-technique, described above, may be called a double PLS-estimation, but involves extraction of relevant information from overlapping factors instead of extraction of relevant information from overlapping manifest variables, as described above. In Table A2 the double PLS-algorithm is presented.

The distribution of regression coefficients estimated by the double PLS-estimation is unknown. By means of bootstrap-techniques, however, we have simulated the precision for the estimated coefficients for the double PLS-estimation. The fundamental motivation of bootstrap is to create the sample distributions from the data at hand when it is difficult, and even impossible, to depend upon any theoretical distribution. According to the central limit theorem, estimators tend to be normally distributed (see for example Efron and Tibshirani, 1993). Patient answers were randomly drawn with replacement from the study data, containing 599 patients. In such a way 1000 new data files, each containing answers from 599 patients, were constructed. We estimated two models by double PLS-estimation technique on each data set, the one in Figure 1 above and one model with an additional latent variable, "The health care centre", with direct path to PSI. Hereby we could capture the variance, and the precision, on the estimated coefficients.

In Table A5 we can see the result. The impact for "The health care centre" that was insignificant when the model was estimated by OLS-technique, becomes significant at every reasonable level when the model is estimated by double PLS-estimation technique. In Table A5 we can see the difference between estimation by OLS, as in ordinary PLS-techniques, and estimation by the double PLS-technique. The relatively small differences in the estimates between the two methods indicate again that the degree of multicollinearity is relatively small. Thus, the insignificant OLS-impact of "The health care centre" is due to the fact that the variance in that variable has little in common with the endogenous variable. On the other hand, the variance in "The health care centre" obviously is relatively unique and has little in common with the other exogenous variables.

5 Analysis of resource use and society cost

A secondary purpose of this study is to analyse the link between patients' HRQL and "Resource use", i.e. we want to analyse if patients who grade their HRQL higher, consume more or less resources from society than patients who grade their HRQL lower. The different types of resources are: "Absence from work" (number of days during last year), "Emergency visits" (number of times during last year), "Visit to a physician" (number of times during last year), "Visit to a nurse" (number of times during last year), "Inhalation steroids" (dose per day, during a normal day), short-term beta 2 stimulator (dose per day, during a normal day) and long-term beta 2 stimulator (dose per day, during a normal day).

Since some of these resources can be operating against each other, we cannot link the aspects of HRQL in the empirical model directly to some aggregate of resources by using PLS-analysis. For example, "Visit to a physician" and "Visit to a nurse" may prevent long-term sickness so that resources are saved for less "Absence from work".

To establish the link between HRQL and "Resource use" with most power, we have linked the separate aspects of HRQL, rather than the factor, to the separate types of resources by using analysis of variance technique. Each aspect of HRQL is put into two groups. Patients who grade the specific aspect of their quality of life: 1) relatively low (scores 1-6) and 2) relatively high (scores 9-10), on a 1 to 10 scale. Patients who graded other scores are deleted to get more power in the test.

The result is shown in Table 3 below. Each value presented in the table implies that improvements in the specific aspect of HRQL lead to reduced resources, i.e. patients who grade the specific aspect of HRQL higher consume fewer resources.

The values presented are the Bonferoni adjusted p-values of the F-tests for each resource type that are significant on the 5% level or lower. Bonferoni adjustments are based on the number of tests directly related to the number of categories and measurement level of a predictor. By Bonferoni adjustments the false-positive error rate (the so-called type 1 error) is better controlled (see Malhotra, 1999). Note that the number of valid cases is different for different questions. Where no links are presented, there is no significant difference in resource use between patients who graded the specific aspect of quality of life high or low.

Table 3. Test by analysis of variance: Reduced perceived problems in the aspects of quality of life result in reduced respective resource use.
(p-values, of 0,05 level or below, on Bonferoni adjusted F-test are presented).

	Types resources							
Aspects of quality of life	Absence from work, days during last year	Emergency visits, number of times during last year	Visit to a physician, number of times during last year	Visit to a nurse, number of times during last year	Days in hospital, number of days during last year	Inhalation steroids, dose per day	Short-term acting beta 2 stimulator, dose per week	Long-term acting beta 2 stimulator, dose per week
Few mental symptoms: I feel worried for side effects in the long run due to my asthma treatment								
Few mental symptoms: I feel physically limited due to my asthma disease			0,001			0,000	0,002	0,005
Few mental symptoms: I feel worried due to my asthma disease				0,046	0,01	0,000	0,000	0,000
Few mental symptoms: I feel socially limited due to my asthma disease	0,05		0,000			0,000	0,000	0,020
Few symptoms at night: I am troubled nightly due to my asthma disease			0,000	0,035		0,000	0,000	0,009

Aspects of quality of life	Types of resources							
	Absence from work, days during last year	Emergency visits, number of times during last year	Visit to a physician, number of times during last year	Visit to a nurse, number of times during last year	Days in hospital, number of days during last year	Inhalation steroids, dose per day	Short-term acting beta 2 stimulator, dose per week	Long-term acting beta 2 stimulator, dose per week
Few other physical symptoms: I have difficulty breathing when I don't take my asthma drug continuously						0,008	0,01	0,003
Few other physical symptoms: I cough when I don't take my asthma drug continuously			0,023			0,000	0,001	0,037
Few other physical symptoms: I feel pressure on my chest when I don't take my asthma drug continuously						0,000	0,002	0,000
Few other physical symptoms: I hear sound from my chest if I don't take my asthma drug continuously						0,000	0,000	0,001

There is only one aspect of HRQL, "I feel worried for side effects in the long run due to my asthma treatment", which does not have any significant link to any type of resource use that we have analysed.

We then look at the cost[3] use for patients treated at asthma health care centres compared to patients treated at ordinary health care centres, and the cost use for patients in the different treatment groups A, B and C, presented in Table 4 and 5 below:

Table 4. Cost for patients treated at asthma health care centres versus patients treated at ordinary health care centres.
p-values on t-test vs. the other group in parenthesis. Average cost during one year in SEK.

Type of cost	**Ordinary health care centres** n=140	**Asthma health care centres** n=503
Drug	4 150 (0,34)	3 763
Visits to physician	1 265 (0,60)	1 212
Visits to nurse	406 (0,01)	306
Emergency visits	1 681 (0,13)	3 460
Days in hospital	1 098 (0,13)	4 391
Absence from work	4 539 (0,35)	7 711
Other treatment	87 (0,08)	171
Total cost	**13 253 (0,07)**	**21 137**

Table 5. Cost for patients in the different treatment groups A, B and C.
Average cost during one year in SEK. *p*-values, on t-test for the level on which we have significant differences, in parenthesis (where p^B means test vs. group B and p^C means test vs. group C).

Type of cost	**Treatment groups** Average cost during one year, SEK		
	A n=204	**B** n=209	**C** n=230
Drug	2 015 (p^B=0,00, p^C=0,00)	3 914 (p^C =0,00)	6 039
Visits to physician	1 114 (p^B=0,37, p^C=0,01)	1 230 (p^C =0,09)	1 383
Visits to nurse	333 (p^B=0,64, p^C=0,001)	350 (p^C =0,005)	459
Emergency visits	1 542 (p^B=0,36, p^C=0,57)	2 859 (p^C =0,44)	1 802
Days in hospital	796 (p^B=0,39, p^C=0,48)	2 560 (p^C =0,84)	2 029
Other treatment	103 (p^B=0,53, p^C=0,57)	75 (p^C =0,21)	133
Absence from work	2 544 (p^B=0,14, p^C=0,28)	7 326 (p^C =0,69)	5 672
Total cost	**8 514** (p^B=0,02, p^C=0,01)	**18 314** (p^C =0,88)	**17 567**

[3] The total cost use per patient is calculated, using cost standards, based on the patients stated doses per day of the drug, number of visits to physicians and nurses during the last year, number of emergency visits and days in hospital during the last year, number of days in absence from work during the last year and other treatment.

From Table 4 we can see that patients treated at asthma health care centres have significantly lower cost for visits to nurses than patients treated at ordinary health care centres. There is also a tendency towards lower cost for absence from work and total cost for patients treated at asthma health care centres, however, due to the few cases in the group treated at asthma health care centres (140 patients) and in particular the large variance in the patient group, we cannot say that these types of costs are significantly lower than for patients treated at ordinary health care centres. The results in Table 5 show for example a significantly lower total cost for patients in treatment group A compared to patients in treatment groups B and C. This can perhaps be explained by the fact that patients within treatment group A have a less severe asthma disease than patients in the other treatment groups. However, we se no significant difference in total cost between treatment groups B and C, which might be due to the large variance within the treatment groups. Furthermore, cost for drug is lower in treatment group A compared to both treatment groups B and C, and costs for visits at physicians and visits at nurses are lower for treatment group A compared to treatment group C. And, costs for drug and visits at nurses are lower for treatment group B compared to treatment group C.

6 Concluding remarks

This study was conducted to determine the most important factors involved in asthma treatment, aimed at making strong recommendations for optimised improvements.

The factors having the largest impact on overall satisfaction, the PSI, are "The drug" and "The physician's manner". These factors also have greatest impact on "Compliance", together with overall satisfaction. The results show that to optimise improvements in overall asthma treatment requires efforts to strive for higher scores for "The physician's manner" and "The physician's medical competence". ""

We also show that patients in the study treated at asthma health care centres are significantly more satisfied with overall asthma treatment, and that they are significantly more compliant and have significantly better asthma control compared to patients treated at an ordinary health care centre.

Furthermore, we see that patients treated with inhaled glucocorticoid + long-term-acting beta-2-agonists in a single inhaler are significantly more compliant than patients treated with glucocorticoid + short-term-acting beta-2-agonists.

Patients having separate inhalers for glucocorticoid + long-term-acting beta-2-agonists rate their HRQL the lowest, while patients having a single inhaler of the same products and patients who inhaled glucocorticoid + short-term-acting beta-2-agonists rate their HRQL rather equally. However, patients using a glucocorticoid + short-term-acting beta-2-agonists probably have a less severe asthma. It is interesting to note that patients having the same products, either in a single inhaler or having two different inhalers, state their HRQL differently. The results also show that improved HRQL implies reduced resources, i.e. patients who grade their HRQL higher consume fewer resources.

Some fields, however, need yet further investigation. For example, our analysis shows that there is a tendency towards less absence from work and lower total resource use for patients treated at asthma health care centres compared to patients treated at ordinary health care centres. Due to the large variance among the patients, perhaps, and few cases in the group treated at asthma health care centres (140 patients), we cannot prove significantly less resource use than for patients treated at ordinary health care centres, using current data.

Similarly, we see a tendency for patients treated with glucocorticoid + long-term-acting beta-2- agonists in separate inhalers to consume more resources in total than patients treated with glucocorticoid + long-term-acting beta-2- agonists in one inhaler. The variance among the investigated patients in those different groups is too large, while we cannot see any significant difference, and further research is needed.

References

Anderson, E. W. and C. Fornell (2000), Foundations of the American Customer Satisfaction Index, *Total Quality Management* 11, 869-882.

Andersson, F., E. Ståhl, P. J. Barnes, C. G. Löfdahl, P. O'Byrne, D. S. Postma, R. A. Pauwels, A. Tattersfield and A. Ullman (2000). For the Formoterol and Corticosteroids Establishing Therapy (FACET) International Study Group. A high dose of inhaled budesonide is more cost-effective than a low dose in moderate asthma. *Journal of Allergy and Clinical Immunology*, 105 (no 1, part 2), 103-4. American Society for Quality (1998), American Customer Satisfaction Index*(ACSI) Methodology Report*, University of Michigan, Ann Arbor.

Berggren, F. and T. Ekström (1999). A cost-effectiveness study comparing the on demand use of formoterol (Oxis®) and terbutaline (Bricanyl®) in patients with moderate asthma. *Eur Resp J*, 14, suppl 30, 341.

Cassel, C., P. Hackl and A. H. Westlund (1999). Robustness of partial least-squares method for estimating latent variable quality structures, Journal of Applied Statistics, 26, 4, 435-446.

Dahlberg, U., T. Ekström, M. Lindberg et al. (1997). Kvalitetsgranskning av astmasjukvård inom primärvården i E-län. Färre symptom och besök med lokalt vårdprogram, *Läkartidnngen*, 94, 3135-3139.

Dijkstra, T. K. (1983). Some comments on maximum likelyhood and partial least squares methods. *Journal of Econometrics*, 22, 67-90.

Eckerlund, I. B. Jönsson, M. Tambour and A. H. Westlund (1997). Change-oriented patient questionnaires – testing a new method at three departments of ophthalmology, *International Journal of Health Care Quality Assurance*, 10, 7, 254-259.

Eckerlund, I. J. Eklöf and J. Nathorst-Böös (2000). Patientönskemål eller kostnadseffektivitet? Prioriteterna påverkas högst påtagligt då en ekonomisk dimension införs. *Läkartidningen*, 97, 5782-5788.

EPSI Technical Committee (1998). *European Customer Satisfaction Index, Foundation and Structure for Harmonized National Pilot Projects*. Report Prepared for the EPSI Steering Committee, October.

Fornell, C. adn J. Cha (1994). Partial Least Squares, In Bagozzi, R.P., (ed), *Advanced Methods in Marketing Research*, Cambridge, Basil Blackwell, 52-78.

Fornell, C. (1992), A National Customer Satisfaction Barometer, the Swedish Experience, *Journal of Marketing*, 56, 6-21.

Fornell, C., M. D. Johnson, E. W. Anderson, J. Cha and B. E. Bryant (1996). The American Customer Satisfaction Index, Nature, Purpose and Findings, *Journal of Marketing*, 60, 7-18.

Gibson, P. G., J. Coughlan, A. J. Wilson, M. Abramson, A. Bauman, M. J. Hensley and E. H. Walters (2000). Self-management education and regular practitioner review for adults with asthma. In *The Cochrane Library*, Issue 3, Oxford.

Gujarati, D. N. (1995). *Basic Econometircs* , 3rd ed. New York, McGraw-Hill.

Henriksson, P., A. H. Westlund, B. Jönsson et al. (1997). Focused Quality Improvement in Health Care: the Quality, Satisfaction, Performance Technique Applied to Internal Medicine, *Quality Management in Health Care*, 5, 4, 76-84.

Johnson, M. D., A. Gustafsson, T. W. Andreassen, L. Lervik and J. Cha (2001). The Evolution and Future of National Customer Satisfaction Index Models, *Journal of Economic Psychology*, 22, 217-245.

Jöreskog, K. G. and H. Wold (1982). The ML and PLS techniques for modelling with latent variables. Systems under indirect observations. Wold, H. New York, North-Holland, 1, 263-270.

Liljas, B. and A. Lahdensuo (1997). Is asthma self-management cost-effective? *Patient Education and Counselling*, 32, 97-104.

Lisspers, K. and E. Österlund Efraimsson (1998). Utvärdering av astmamottagningen på vårdcentralen Tisken 1996 och 1997. Tiskens Vårdcentral, Falun.

Lohmöller, J. B. (1989). Latent Variable Path Modeling with Partial Least Squares, Heidelberg, Physica-Verlag.

National Board of Health and Welfare (2003). Socialstyrelsens riktlinjer för förebyggande diagnostik, behandling och rehabilitering av astma och KOL - Remissversion, Stockholm.

Lorentzson, S. (1995). Ekonomiska effekter av bättre vårdformer. *Svensk Farmacevtisk Tidskrift*, 99, (suppl), 32-36.

Ramanathan, R. (1998). *Introductory Econometric with applications* , 4th ed, New York, Harcourt Brace.

The Swedish Council on Technology Assessment in Health Care (2000). Behandling av astma och KOL. En systematisk kunskapssammanställning. Stockholm.

Westlund, A. H. (1994). "Nöjda Patienter? Åtgärder prioriteras med ny metod. In "*Sprida Tankar om vården*. SPRI.

Westlund, A. H. I. Eckerlund, M. Tambour et al. (1998). Åtgärdsinriktade Patientenkäter. Ny lovande metod prövad vid tre ögonkliniker. *Läkartidningen*, 93, 2078-2081.

Wold, H. (1966). Nonlinear estimation by iterative least squares procedures. In David F.N. (ed), *Festschrift for J. Neyman: Research papers in statistics*, London, Wiley, 411-444.

Wold, H. (1982). *Soft Modeling: The Basic Design and Some Extensions*, in Jöreskog, K.G. and Wold, H. (ed), System under Indirect Observations: Causality, Structure, Prediction, Amsterdam, North-Holland, 2, 1-54.

Zetterström, O., R. Buhl, H. Mellem and F. Andersson (2002). *The whole story: treatment outcomes with Symbicort*. Respiratory Medicine. Feb. 96, Suppl A, 29-35.

Appendix

Table A1. Factors and statements in the questionnaire. (Factors are in bold typeface while each statement is marked with a bullet)

The physician's medical competencs

- that your physician/s is/are competent within the field of asthma diseases
- the physicians ability to explain in a way that you understand
- the physician's ability to give you sufficient information about asthma as a disease
- the physician's ability to give you sufficient information of how you can live with your asthma in your daily life
- the physician's ability to give you sufficient information on how to dose your asthma medicine
- the physician's ability to give you sufficient information on how to use your asthma medicine
- the physician's ability to give you sufficient information on the side effects of your asthma medicine

The physician's manner

- that the physician gives you time
- that the physician listens to you
- that the physician is considerate
- that the physician shows commitment to you as a patient
- that you feel confident in your physician

The nurse's medical competence

- that the nurse/s is/are competent within the field of asthma diseases
- the nurse's ability to explain in a way that you understand
- the nurse's ability to give you sufficient information about asthma as a disease
- the nurse's ability to give you sufficient information of how you can live with your asthma in your daily life
- the nurse's ability to give you sufficient information of how to dose your asthma medicine
- the nurse's ability to gives you sufficient information on how to use your asthma medicine
- the nurse's ability to give you sufficient information on the side effects of your asthma medicine

The nurse's manner

- that the nurse gives you time
- that the nurse listens to you
- that the nurse is considerate
- that the nurse shows commitment to you as a patient
- that you feel confident in your nurse

Availability of the health centre

- that it is easy to book an appointment by telephone
- that you can book a time for treatment of you asthma that is suitable for you
- that you can reach the physician by telephone concerning your asthma
- that you can reach the nurse by telephone concerning your asthma

The health centre

- that it is pleasant
- that it is light and airy
- that it is clean and well-kept
- that it is a place where you can talk with the staff without interruption

The other staff's manner

- that other staff are available
- that the other staff give you time
- that the other staff assist you correctly

Table A1. Continuation. Statements in the questionnaire

The effect of the drug

- I can feel immediate effect when I have taken my asthma drug
- my asthma drug makes it easier for me to control my asthma treatment

The side effects of the drug

- I don't have any side effects of the asthma drug

The ease of usage of the drug

- that it is easy to use
- that it is easy to have about you

Absence of price elasticity of the drug

- I have always picked up my asthma drug regardless of my economical situation

The drug

- I am generally satisfied with my asthma drug

Patient satisfaction index

- overall how satisfied are you with your asthma treatment
- how well does your asthma treatment meet your expectations
- imagine an asthma treatment that is perfect in every sense. How close is your current asthma treatment to this ideal

Compliance

- I follow my physician's recommendations about the dose of the asthma drug carefully

Asthma control

- I feel that my asthma disease is under control
- I feel that it is I who manage my asthma treatment

Few mental symptoms

- I feel worried for side effects in the long run due to my asthma treatment
- I feel physically limited due to my asthma disease
- I feel worried due to my asthma disease
- I feel socially limited due to my asthma disease

Few symptoms at night

- I am troubled nightly due to my asthma disease

Few other physical symptoms

- I have difficulty breathing if I don't take my asthma drug regularly
- I get cough if I don't take my asthma drug regularly
- I can hear sound from my chest if I don't take my asthma drug regularly

Note: All questions are scored on a 1 to 10 scale by the patients, the higher the score the better.

Table A1. Continuation. Background variables

Information about whether the patient visits an asthma health care centre or an ordinary health care centre.	**Number of visits to physicians during one year?**
Gender?	**How many nurses do you see for your asthma?**
Age?	**Number of visits to nurses during one year?**
Type of education?	**Questions about type of drug, how often the drug is used during one day and how many doses on each occasion.**
Life situation? (Student, working, retired etc.)	**Other treatments than the prescribed drug?** (i.e. health food etc.)
When did you first have your asthma diagnosed?	**Cost for this other treatment during one year?**
How long before your asthma diagnosis do you think you have had symptoms due to your asthma disease?	**How many emergency visits during one year due to your asthma disease?**
Number of days with asthma symptoms during a normal month?	**How many days in hospital during one year due to your asthma disease?**
How many physicians do you see for your asthma?	**(If applicable) How many days absence from your work during one year due to your asthma disease?**

Table A2. The double PLS algorithm.

For the first phantom variable	**Scaling** $X_1 = X - \overline{X}, \quad y_1 = y - \overline{y}$
	Optimising $\underset{w_1}{Max} \quad w_1' X' y_1 \quad w_1' w_1 = 1$
	Phantom variable $\xi_1 = X_1 w_1$
	Impact for phantom variable on y $c_1 = \left(\xi_1' \xi_k\right)^{-1} \xi_1' y_1$
	Impact for quality component X on y $b_1 = w_1 c_1$
For the following phantom variable	**Scaling** $X_k = X_{k-1} - \lambda_{k-1} \xi_{k-1}, \quad y_k = y_{k-1} - \xi_{k-1} b_{k-1}$
	Optimising $\underset{w_1}{Max} \quad w_k' X_k' y_k \quad s.t. \; w_k' w_k = 1$
	Phantom variable $\xi_k = X_k w_k$
	Impact for phantom variable on y $c_k = \left(\xi_{k1}' \xi_k\right)^{-1} \xi_k' y_k$
	Impact for quality component X on y $b_k = w_k c_k$
	Total impact $\gamma = \sum_{k=1}^{Npv} b_k$

Table A3. First two eigenvalues of the factors within the model, reliability (Cronbach's alpha) and average variance extracted

	First eigenvalue	Second eigenvalue	Reliability (Cronbach's alpha)	AVE	R^2
The physician's medical competence	5.3	0.5	0.95	0.76	
The physician's manner	4.6	0.2	0.98	0.91	0.56
The nurse's medical competence	5.8	0.5	0.96	0.82	0.62
The nurse's manner	4.6	0.2	0.98	0.92	
Availability of the health centre	2.8	0.5	0.86	0.71	
The health centre	2.8	0.6	0.86	0.71	
The other staff's manner	2.7	0.2	-	0.91	
The effect of the drug	1.4	0.6	0.59	0.70	
The side effects of the drug	-	-	0.76	1	
The ease of usage of the drug	1.6	0.4	-	0.81	
Absence of price elasticity of the drug	-	-	-	1	
The drug	-	-	0.93	1	0.44
Patient satisfaction index	2.6	0.2	-	0.88	0.57
Compliance	-	-	0.83	1	0.15
Asthma control	1.7	0.3	0.77	0.85	0.24
Few mental symptoms	2.4	0.7	0.75	0.60	0.17
Few symptoms at night	-	-	-	1	0.18

Table A4. Collinearity diagnostics for empirical model, condition numbers.

	Largest condition number
The drug	20.32
Patient satisfaction index	17.44
Compliance	14.40
Asthma control	11.00
Few symptoms at night	8.65

Table A5. Estimated coefficients by ordinary PLS- (OLS) and double PLS-technique (dPLS). *p*-values in parenthesis.

Independent latent variables	Dependent latent variables			
	The physician's manner	The nurse's manner	The drug	Patient satisfaction index
The physician's medical competence	OLS 0.81(0.00)			
The physician's manner				OLS 0.36(0.00) dPLS 0.32(0.00)
The nurse's medical competence		OLS 0.67(0.00)		
The nurse's manner				OLS 0.2(0.00) dPLS 0.14(0.00)
Availability of the health centre				OLS 0.08(0.01) dPLS 0.086(0.00)
The health centre				dPLS 0.0981(0.00)
The effect of the drug			OLS 0.38(0.00) dPLS 0.34(0.00)	
The side effects of the drug			OLS 0.12(0.00) dPLS 0.12(0.00	
The ease of usage of the drug			OLS 0.4(0.00) dPLS 0.41.0(0.00)	
Absence of price elasticity of the drug			OLS 0.1(0.00) dPLS 0.1(0.00)	
The drug				OLS 0.34(0.00) dPLS 0.35(0.00)

Table A5. Continuation. Estimated coefficients by ordinary PLS- (OLS) and double PLS-technique (dPLS). *p*-values in parenthesis.

Independent latent variables	Dependent latent variables			
	Compliance	Asthma control	Few mental symptoms	Few symptoms at night
The physician's manner	OLS 0.12(0.02) dPLS 0.12(0.02)			
The drug	OLS 0.22(0.00) dPLS 0.22(0.00)	OLS0.12(0.00)		
Patient satisfaction index	OLS 0.16(0.01) dPLS 0.17(0.01)	OLS 0.41(0.00)		
Asthma control			OLS 0.37(0.00)	OLS 0.15(0.00)
Few mental symptoms				OLS 0.33(0.00)

Willingness to pay for anticoagulant treatment: A comparison of conjoint analysis and contingent valuation

Anna Ovanfors
Department of Economic Statistics, Stockholm School of Economics
Box 6501, SE-113 83 Stockholm, Sweden
October 14, 2004

Abstract

The aim of this study is to estimate the willingness to pay for improved oral anticoagulant treatment, and to compare two different methods for estimating the willingness to pay (conjoint analysis and dichotomous choice contingent valuation).

682 patients, currently receiving the standard anticoagulant treatment (warfarin), are included in the study. The patients are asked to value an oral drug with fewer side effects and fewer limitations in daily life, in comparison to a new preparation of which they have no experience.

To test for a scope effect, the limitations in daily life are varied between "few limitations in daily life" and "many limitations in daily life" in two different sub samples. In the conjoint analysis the patients rank 18 treatment concepts including the warfarin-like treatment and the two treatment alternatives in the contingent valuation study.

The results suggest that patients are willing to pay a substantial amount for the improved anticoagulant treatment, and that the dichotomous choice contingent valuation method and conjoint analysis produce relatively similar results. The estimated monthly willingness to pay for the improved treatment is about 800 SEK according to the conjoint analysis and 700 SEK according to the contingent valuation study. The estimated scope effect is significantly smaller with contingent valuation than with conjoint analysis.

Key words. Oral anticoagulant (OA) treatment, conjoint, contingent valuation, treatment preferences, warfarin treatment, willingness to pay (WTP).

1 Introduction

The most common cause of death in industrialized countries is cardio-vascular disease, and in Sweden it causes almost 50% of all deaths among men and women (National Board of Health and Welfare – Health and Disease Statistics 2001:7).

Patients with chronic atrial fibrillation (a common heart rhythm disorder) are 5-6 times more likely to have a heart attack with symptoms (Petersen et al. 1990). It is estimated that around 6,000 strokes per year are caused by atrial fibrillation (se Asplund et al. 1995; and Eriksson et al. 1987), from the point of view of the prevalence of the risk factor.

Treatment with oral anticoagulants (OA treatment), most many warfarin, protects a large number of patients every year from thromboembolism when they have venous thrombosis, atrial fibrillation or mechanical cardiac valve (Själander et al. 2003). Since the OA-treatment increases the risk of haemorrhage causing sudden death, hospitalisation or invalidity, the treatment demands specialist care (Johnsson, 1999). Severe haemorrhaging places high demand on health care system and the average cost is high (Terent et al. 1998).

The use of warfarin is increasing, and there are also increased demands from both patients and society for the treatment to be easy to use, safe and that it should be given on strict indications (Johnsson, 1999). Warfarin treatment is a balance between high enough doses to prevent ischemic stroke, but not too high to increase the risk of haemorrhage. Warfarin also interacts with food and other drugs. These characteristics lead to restrictions in the patients' daily living.

A substance, which we choose to refer to as H, is a new oral anticoagulant that has been evaluated in clinical studies for the indication preventing thrombi embolism in patients with atrial fibrillation (see for example Eriksson, 2002; Olsson, 2003; and Wallentin et al. 2003). Studies show that H is as good as warfarin in preventing stroke and is associated with reduced risk of haemorrhage compared to warfarin. (Eriksson, 2002; Olsson, 2003; Wallentin et al. 2003). However, in addition H has another treatment profile: the dosing regime is different, no need for anticoagulation monitoring, and the limitations in daily life are less since there is no interaction with food and low potential for interaction with other drugs. The effect of H still seems to be as good as, and in many times better than, the treatment with warfarin (Eriksson, 2002; Olsson, 2003; and Wallentin et al. 2003). However, how the patients value these treatment characteristics is unknown.

The first purpose of this study is to analyse preferences for different OA treatments among patients treated with warfarin. The second purpose is to compare conjoint analysis and contingent valuation when estimating the patients' mean willingness to pay for alternative treatment. Willingness to pay estimates provide information of how patients valuate a treatment. Therefore, willingness to pay estimates are valuable when it comes to deciding of whether the drug should be included in health insurance (see Telser and Zweifel, 2002).

The plan of the paper is as follows. Firstly, we will discuss the data collection and the design of the study. Secondly, we give a short summary of the conjoint- and the contingent valuation method chosen for this study. Thirdly, the results of the analysis will be discussed, and finally we will conclude the paper.

2 Data collection

2.1 Comprehensive design of the study

This study comprised 682 Swedish patients diagnosed as having atrial fibrillation, and was performed as a survey-based multi-center preference study on patients' own experiences of their overall OA treatment. However, each patient was informed of different alternative treatments with drugs before answering the survey. Between 23 April and 23 June, 29 participating general practitioners and 16 anti coagulation clinics collected the study data consecutively during standard warfarin treatment checks. No drugs were used in the study.

The risk of including only patients with special qualities, i.e. a risk of getting a systematically skewed sample, is minimized in a consecutive procedure since it is not likely that patients with special characteristics are coming in a sequence. Given our study design the non-response rate and the amount of incomplete answers are also minimized. 678 patients were used in the analyses and the non-response rate was thus 1%, which must be considered low in these types of studies (see for example; Ryan and Hughes, 1997; Aristides et al. 2004).

Patients were asked to complete a questionnaire, then answer a question according to the contingent valuation method and finally to rank treatment concepts according to the conjoint method (see Figure A1, Table 1 and Table A2). In the questionnaire, patients reported demographic information including age, gender, exercise, education and economic situation. The patients filled in the first question regarding their warfarin treatment with assistance from a physician or nurse. The rest of the questionnaire, the contingent valuation question and the ranking exercise were completed by the patients themselves.

The general information broke down the warfarin treatment aspects into attributes and levels that were described in detail, without any substance name or product name. The attributes and levels were also presented on 18 cards as different treatment concepts (see Table 1). Some treatment concepts were similar to warfarin treatment (see W1, W2 or W3 in Table 5) or to treatment with the drug H (see X1 in Table 5); other treatment concepts were hypothetical, unlike warfarin treatment and H treatment.

Local ethics committees approved the study, and all patients gave their written informed consent.

2.2 The study population

The study involved patients, regardless of age, who were receiving warfarin treatment.

Inclusion criteria

- Patients diagnosed with chronically atrial fibrillation
- Patients having received warfarin treatment for at least 6 months

Exclusion criteria

- Individuals insufficiently able to read and write in Swedish.
- Individuals with reduced cognitive ability.
- Individuals who were part of another anticoagulant treatment study at the same time.
-

3 Methodology

3.1 Choice of analysis technique

The first purpose of this study is to analyse preferences for different OA treatments among patients receiving warfarin treatment.
The second purpose is to compare conjoint analysis and contingent valuation when estimating the patients' mean willingness to pay for alternative treatment. Willingness to pay estimates provide information on how the patients' value treatment. Therefore, willingness to pay estimates are necessary when it comes to deciding whether the drug should be included in health insurance (see Telser and Zweifel, 2002).

The first purpose of the study was to analyse preferences for different anticoagulant treatments for patients receiving warfarin treatment. For this purpose, a patient-specific ranking of hypothetical treatment alternatives in accordance with the conjoint methodology was performed (see Table 1). The attributes and the attribute levels were based upon extensive qualitative interviews with physicians and experts. The selected attributes were "Number of INR checks" (which are checks at physician testing the blood for correct warfarin dose), "Side effects", "Dose", "Limitations in daily life" and "Monthly cost". The individual patient's utility is not directly observable in conjoint analysis, but can be derived when the patient ranks a number of experimentally designed treatment combinations with varying attribute levels. We find out how the attributes and the attribute levels are preferred compared to each other, by introducing new hypothetical treatment concepts of which the patient has no experience.

Conjoint analysis was originally developed for market research to study consumer preferences, and the method investigates the relative importance of group attributes, e.g. products with certain properties or more abstract concepts such as treatment procedures (see for example Green and Srinivasan, 1990; Gustafsson et al. 2001; Cattin and Wittink, 1982). It has been applied to various aspects of health care (for reviews see Ryan and Hughes, 1997; Johnson et al. 1998; Aristides et al. 2004; and Telser and Zweifel, 2002).

The second purpose of this study was to estimate the patients' mean willingness to pay for alternative treatment comparing conjoint analysis and contingent valuation. Telser and Zweifel, 2002 discuss advantages and disadvantages of the two most commonly used methods used for this purpose, the conjoint approach and contingent valuation. In our analysis, however, we compare the two methods and test them against each other.

One of the issues of concern in a conjoint study is choosing the appropriate research design. Many the task is to choose between choice-based conjoint analysis and traditional design (e.g. rating, ranking) approaches, both methods are commonly used (see for example Ryan and Hughes, 1997; Aristides et al. 2004; and Johnson et al. 1998). In the study by Telser and Zweifel (2002), the patients were asked to state whether or not they would buy the offered product.

In our study we have chosen the ranking method for a number of reasons. Firstly, choice-based conjoint data is many collected using interactive media (e.g. touch pads, computer-assisted questionnaire), which in our case dealing with elderly persons might cause non-response due to the technology employed. Furthermore, choice-based conjoint analysis produces relatively precise results when there are few attributes and their interaction is of

concern (Green and Srinivasan, 1990). The preference study aimed at ranking a large number of treatment concepts on many attributes, which is why the choice-based conjoint approach would be less suitable in this study.

Another advantage of including both conjoint analysis and contingent valuation in the study is that it enables us to estimate the magnitude of starting point bias in the conjoint analysis by means of the contingent valuation offer (i.e. the anchoring effect towards the bids in the contingent valuation question). In the contingent valuation question (see Appendix Figure A1), every patient was given only one price alternative and one treatment concept. However, the patients were neither given the same price alternative nor the same treatment concept. The patients answered whether they wanted to pay the given amount of money out of their own pocket for the treatment. Thus, we divide the patients into 4 groups according to the bids they were offered, with the purpose of seeing if the patients ranked the treatment concepts differently in the conjoint analysis due to the offer they were given in the contingent valuation question.

3.2 Conjoint analysis

In the conjoint analysis, several characteristics or attributes of the treatment are identified (see Table 1 below), and for each attribute a range of possible values, or attribute levels, are defined. In this way it is possible to create a number of "hypothetical" treatment concepts, each with different levels for various attributes.

Table 1. Conjoint design. Treatment concepts are constructed using combinations of different attribute levels.

Attributes/ treatment characteristics	***Attribute levels***
Number of checks	C_1 = 18 checks per year C_2 = 12 checks per year C_3 = One blood-test every month for the first six months, after that no more checks
Side effects	S_1 = Your drug is unlikely to interact with many other drugs including pain killers, which can result in increased risk for minor bleeding like nose bleeds, bleeding from your gum and bruising. S_2 = Your drug may interact with other drugs, which decreases the risk for minor bleeding like nose bleeds, bleeding from your gum and bruising.
Dose	D_1 = One pill twice a day, and the number of pills never changes. D_2 = Pills are taken once a day, but the number of pills can vary during the week.
Limitations in daily life	L_1 = Many L_2 = Some L_3 = Few
Monthly cost for the treatment, SEK	P_1 = 975 P_2 = 650 P_3 = 325 P_4 = 30

The number of possible treatment concepts that could be constructed from the attribute levels in Table 1 is $3^2 \cdot 2^2 \cdot 4 = 144$. In practice, it is impossible for an individual to rank all these treatment concepts. Instead of presenting the respondent with all alternatives, it is possible to choose only a fraction of alternatives, i.e. a fractional factorial design. In this study we have

generated orthogonal main-effect fractional factorial designs, and thus the number of treatment concepts presented to the patients was reduced to 18 (see Table 5 below for examples). This was done using the SPSS Orthoplan Procedure (se SPSS, 2004).

Sixteen of the 18 cards were used to calculate preferences, and the remaining 2, the so-called "holdouts", were used to test the reliability of the results. Holdouts are cards that are ranked by the patients but that are kept out of the estimation. The ranking of holdouts is compared to the expected utility for the holdouts based on the estimated function. Kendall's tau is an established measure of validity in conjoint studies (see for example Wallsten, 1967; Kendall, 1970).

In this study we have chosen Ordinary Least Squares (OLS) as an estimation approach in the conjoint analysis, because OLS estimation is easy to perform and interpret and because it seems to be as good as other more complicated and resource-intensive approaches (Gustafsson et al. 2001).

The patients' rankings are converted into utility estimates and the parameters in each patient's individual utility functions are estimated. For more details see Ryan, (1996) and Gustafsson et al. (2001). These individual parameter estimates are the base for the estimation of the aggregate utility function. It is assumed that the difference between the preference orderings between any two treatment concepts next to each other in the ranking exercise yields the same difference in utility all the time. Thus, the difference in utility between treatment concepts with rankings 1 and 2 is the same as the difference in utility between the concepts with rankings 3 and 4. It is also assumed that the overall utility can be explained as an additive decomposition of the utilities of different elements as described below:

$$Y_k = \sum_{j=1}^{J} \sum_{l=1}^{L} \beta_{jl} I(X_j = x_{jl}) + \mu, \quad \text{for } k = 1,...,K \text{ and } \forall j \sum_{l=1}^{L} \beta_{jl} = 0 \tag{1}$$

where Y_k denotes the total utility of concept k, X_j denote the attributes, x_{jl} are the attribute levels and the parameter estimates, β_{jl}, are the part-worth utilities. $I(.)$ is the indicator function. Preference parameters were estimated by means of OLS for each patient, which means that preferences can be expressed for each patient in the survey, or for subgroups of patients with similar clinical status, socio-demographic profile, or preferences.

All attributes were considered discrete in the estimation. A marginal utility of money (i.e. the utility/SEK) is calculated using the aggregated utility function in Table 2 below (see for example Lancaster, 1971 and Aristide, 2004). Mean willingness to pay can then be estimated for all possible treatment combinations within the price range in the study.

It should, however, be noted that we do not have equidistance between the steps in "Monthly cost". We wanted the lowest level of "Monthly cost" to be 30 SEK/month, since this was considered the average monthly cost of warfarin treatment in Sweden at the time of the study. With this price level included in the conjoint study, we were able to perform a simulation study and compare different treatment concepts with each other with realistic price levels. The highest level, on the other hand, had to be reasonably low so as not to cause protest answers and was set as 975 SEK/month. Equidistance between the levels in "Monthly cost" would have been: 30, 345, 660 and 975 SEK/month. However, monthly costs like 345 and 660 are

not psychologically good choices since they do not seem natural in people's minds, and for this reason we choose the levels: 30, 325, 650 and 975 SEK/month respectively instead.

We have chosen not to exclude patients that displayes price insensitivity (so called irrational traders), since this approach is conservative (see Aristide, 2004). Instead we have chosen to verify the mean willingness to pay estimated by the conjoint method with mean willingness to pay estimated by the contingent valuation method, as described below.

Test for starting point bias in the conjoint analysis

There is a potential source of so-called starting-point bias, when using conjoint analysis to estimate the mean willingness to pay. The intuition behind starting point bias is that the initial starting values affect the final answer. Thus, in the conjoint analysis the rankings might be biased towards the initial value. The possible existence of starting-point bias may be interpreted as unstable responders' preferences (Randall et al. 1978; Brookshire et al. 1980; Thayer, 1981; and Boyle et al. 1985).

This study is designed to allow for tests of the presence of certain starting point bias. By means of a contingent valuation question (see Appendix Figure A1), the respondents are randomized to different starting-bids (monthly cost in SEK): 100, 500, 900 and 1300 respectively. Thus the presence of starting point bias towards these bids in the contingent valuation question can be examined by testing for differences in patients' ranking of treatment concepts, grouped by starting bids.

Mean willingness to pay estimated by the contingent valuation method and the conjoint method can be said to be equally relevant since in the contingent valuation method we have no starting point bias but little information while in the conjoint method we have the reversed situation.

3.3 Contingent valuation

While the conjoint method allows us to estimate indirect mean willingness to pay by means of the estimated marginal utility of money, the contingent valuation question can be used to analyse the patients' direct mean willingness to pay. In the binary contingent valuation question, we ask whether the patients' are willing to pay a certain amount of money out of their own pocket an alternative treatment (see Coursey et al. 1987; Hanemann, 1991; and National Oceanic and Atmospheric Administration, 1993 for discussion). Each patient was offered only one bid to accept or reject. The limitation with this discrete estimation of direct mean willingness to pay, however, is that large samples are needed (Herriges and Shogren, 1996).

In the contingent valuation-question we have used the following bid vector: 100, 500, 900, 1300 SEK per month. In the contingent valuation study we compare the presented alternative, a H alternative, with the patients' treatment today. We assume that the bid alternative 30 SEK per month would have been trivial, i.e. 100% of the patients would have accepted this bid. Furthermore, due to the limited group of patients included in the study, we could not divide them into too many groups and still have reasonable power in the results, therefore the choice was to exclude the bid 30 SEK per month.

In Sweden there is cost minimization on purchases of drugs and treatment by physicians or nurses. The contingent valuation question, however, has to be presented in the context of the patients' willingness to pay out of their own pocket while we want the patients to value their

own utility rather than the society's (see for example Johannesson and Jönsson, 1991; Johannesson, 1993, 1996). In this way the study will be relevant even if the government changes the conditions for cost minimization.

Usually some assumptions have to be made when mean willingness to pay is estimated by means of the contingent valuation method. Thus we take such background variables as degree of illness, income, gender and age into account (see for example Johannesson and Jönsson, 1991; Johannesson, 1993, 1996; and McFadden, 1994).

For mean willingness to pay estimation in the contingent valuation analysis we have chosen one parametric and one non-parametric approach:

Willingness to pay estimated by logistic regression

In this case a negative willingness to pay can be ruled out. Then, using the logistic cumulative distribution function, we have that expected willingness to pay, *E(WTP)*, is given by:

$$E(WTP) = \int_0^{\infty} \frac{1}{1+e^{-\Delta B}} dB = -\frac{1}{\beta} ln\left[1+e^{a}\right] \qquad (2)$$

where B is bid vector, β is the marginal utility of income and a is a constant (see Johansson, 1995 for discussion).

Willingness to pay estimated by Kriström's method

In Kriström (1990), a non-parametric estimation of the mean willingness to pay is proposed, which is based on the proportion of accepts for the contingent valuation offer. The advantage of this method is its simplicity and that it makes no distributional assumptions. If Y_j is the number of accepts for the bids respectively and N_j is the number of non-accepts for the bids we have that:

$$\hat{\pi} = \sum_{j=1}^{m} \frac{Y_j}{Y_j + N_j}$$

where $\hat{\pi}$ is the distribution free Maximum Likelihood estimator of the probability of accepts in the contingent valuation offer if the sequence is monotonic and non-increasing (see Kriström, 1990). This method makes the assumption that $\hat{\pi} = 1$ when the bid equals 0, and $\hat{\pi} = 0$ for the highest bid. Furthermore Kriström makes the assumption that the true distribution function is linear.

4 Results

68% of the patients in the study were men. The average age of the patients was 72, which is in line with what you would expect. Only 8% were below the age of 60 years. 19% took warfarin for secondary preventive cause, and of them 84% had a stroke/tia and 16% had a periphery embolism. The average amount of warfarin checks among the patients was 19 per year.

12% of the patients were gainfully employed, of which 62% worked full time. Of the patients in gainful employment, 8% were absent from work during the check. 22% of the patients had a relative or a friend with them at he check, and of those 16% were absent from work. 78% took less than an hour off wor for the check, 19% took between one and two hours off work and 3% took more than two hours off work.

The first part of this section studies preferences for different anti-coagulant treatments for patients that were treated with warfarin. In the second part of this section we compare conjoint analysis and the contingent valuation method when the patients' mean willingness to pay is estimated.

4.1 Conjoint analysis

The estimated aggregated utility function was as follows (standard error in parenthesis):

Table 2. Estimated aggregated utility function.

Attribute levels	**Utility estimates and standard errors**
C_1	-0.5 (0.096)
C_2	0.34 (0.11)
C_3	0.16 (0.11)
S_1	-1.44 (0.072)
S_2	1.44 (0.072)
D_1	-0.037 (0.072)
D_2	0.037 (0.072)
L_1	-1.24 (0.096)
L_2	0.31 (0.11)
L_3	0.93 (0.11)
P_1	-2.64 (0.12)
P_2	-0.76 (0.12)
P_3	1.14 (0.12)
P_4	2.26 (0.12)
Constant	8.93 (0.08)
Marginal utility of money	$= -\frac{P_4 - P_1}{975 - 30} = -0.0052$
Pearson's R	0.998
Kendall's τ	0.950
Kendall's τ for 2 holdouts	1.00

Where U is the aggregated utility, C_n, S_n, D_n, L_n and P_n are the attribute levels in accordance with Appendix Table 1. A constant term is included in the model, but for rank data the size will vary as a function of the number of cards, and thus is of no interpretative importance.

The parameter estimates, the part-worth utilities, indicate how each attribute level relates to preference. Positive and negative values indicate how utility is changed when moving from one attribute level to another. For each factor, the utilities are normalized so that they sum to 0. Two times the standard error provides an approximate 95% confidence band of the utility estimate. Thus, from the aggregate utility function in Table 2, we find no significant difference between the utility estimates for levels C_2 = 12 checks per year and level C_3 = One blood test every month the first six months, after that no more checks. The same holds for the difference between levels D_1 = One pill twice a day, and the number of pills never changes and level D_2 = The pills are taken once a day, but the number of pills can vary during the week. The other differences are significant.

The importance values in Table 3 below measure how important each attribute is to the preference ordering. Importance values are computed by dividing the range of utility values for an attribute by the sum of these ranges across all attributes, and multiplying then by 100. Since the importance values sum to 100, they can be treated as percentages.

In conjoint analysis we can make a statistical test of homogeneity in the patients' preference structure by dividing the patients according to background factors like gender, age and economic situation. Some of the results are shown in Table 3 below.

Table 3. The attributes' relative importance on total utility for three analysed groups.

Attributes/treatment characteristics	***Importance, in percent***		
	All patients	Patients with selfgraded private economy that is normal	Patients with selfgraded private economy good or very good
Number of checks	8	8	7
Side effects	26	25	28
Dose	1	1	0
Limitations in daily life	20	16	23
Monthly cost for the treatment, SEK	45	50	42

From Table 3 we can see that the attributes' relative importance is only slightly affected by the economic situation. Individual preference profiles have been calculated for each individual with complete ranking data for the 18 treatment concepts. Analysis shows that preference profiles are homogenous with respect to; age, sex, health factors, experiences of different treatment alternatives and economic background.

4.1.1 Validity of the conjoint study

Pearson's R is the standard product moment correlation coefficient and Kendall's tau is an association measure based on the patient's ranking. Both measures are bounded between zero and unity, indicating a perfect relationship between the patients' rankings and the model predictions. Kendall's tau can be calculated for a group of patients as well as for individuals and for the holdouts (see Wallsten, 1967; Kendall, 1970).

We have used only two holdout cards, meaning that Kendall's tau might be affected if one patient's rank order is reversed. If so, the interpretation might be that the group of patients is not homogeneous and that they could be divided into subgroups, or even a more complex model could apply to the data or to a group of patients. The former alternative will be analyzed by means of a cluster analysis.

As the patients seems to be relatively homogenous with respect to demography and socioeconomic situation, we want to see if we can identify natural groups by means of cluster analysis. In conjoint analyses, utility functions are estimated for each patient and utility values are estimated for all attribute levels. The utility estimates are used as cluster variables. The basic criterion used for this is distance, in the sense that patients close together should fall into a cluster that would be relatively homogenous but different from those contained in other clusters. When clustering is successful, the results suggest separate segments within the overall set of data (see for example Gustafsson et al. 2001; Jain et al. 1979; and Green and Srinivasan, 1990).

Chi-square tests have been used to identify significant differences. In Table 4 below we can see the results for three identified clusters.

Table 4. Three defined clusters among the patients. The figures are the attributes' relative importance for total utility.

	Importance, in percent		
Attributes/treatment concept	***Cluster groups***		
	The hypothetical H patient	The patient who likes checks	The price sensitive patient
Number of checks	15	51	9
Side effects	39	16	16
Dose	2	12	0
Limitations in daily life	36	3	11
Monthly cost for the treatment, SEK	7	18	64

From Table 4 above we can see that the hypothetical H patient prefers a drug that results in fewer limitations in daily life and fewer side effects and shows no preference for number of checks and doses. Analyses show that the hypothetical H patient works more, is better off and is somewhat younger than the average patient in the other two segments.

4.1.2 Simulation study

The attribute utility estimates from the conjoint analysis are applied to the attribute combination present in the simulation. This generates a utility estimate for the simulation. Since patients need not see simulation cards, simulations can be designed after the patient data have been collected.

There are a number of theoretical models concerning how to move from the utilities of a set of objects to predictions about preference and choice behaviour. In the Bradley, Terry, Luce (see for example Gustafsson et al. 2001) model the probability of choosing one alternative from a set is equal to the utility of the alternative divided by the sum of the utilities of all alternatives in the set. It assumes all alternatives are positive. It is a probabilistic model since it assigns a probability of choice to each alternative. Under the Bradley, Terry, Luce model, the probability of choosing product i from a set of G products is:

$$P_i = \frac{U_i}{\sum_{i=1}^{G} U_i}$$

where U_i is the estimated utility of the i^{th} treatment concept.

In Table 5 two H alternatives and three warfarin alternatives are described. We consider these treatment concepts to be the most interesting for comparison since they are realistic, and since they reflect the alternatives in the contingent valuation-offer, see Table 7a and 7b below.

In Table 6 we can see the result from the simulation study, showing which treatment concepts are most preferred by patients on average. In Table 6 we consentrate on realistic warfarin- and H alternatives, X1 (at varied monthly cost), W1, W2 and W3 (at 30 SEK/month), described below in Table 5.

Table 5. H and warfarin treatment concepts constructed from the attributes and the attribute levels described in Table 1.

Attributes	*Treatment concepts*				
	Realistic H alternative	**H alternative for comparison**	**Realistic warfarin alternative**		
	X1	**X2**	**W1**	**W2**	**W3**
Number of checks	One blood-test every month the first six months, after that no more checks	One blood-test every month the first six months, after that no more checks	18 checks per year	12 checks per year	18 checks per year
Side effects	Your drug is unlikely to interact...	Your drug unlikely to interact...	Your drug may interact...	Your drug may interact...	Your may interact...
Dose	One pill twice a day, and the number of pills never changes.	One pill twice a day, and the number of pills never changes.	The pills are taken once a day, but the number of pills can vary during the week.	The pills are taken once a day, but the number of pills can vary during the week.	The pills are taken once a day, but the number of pills can vary during the week.
Limitations in daily life	Few	Many	Many	Some	Some
Monthly cost, SEK	Varied	Varied	30	30	30

Table 6. Most preferred treatment concepts by patients on average. Only the realistic H alternative and the warfarin alternatives at 30 SEK/month are shown.

	X1[*)]				**W2**	**W3**
Monthly cost, SEK	30	325	650	975	30	30
Ranking	1	2	3	6	5	4

[*)] X1, at all price alternatives, is preferred to W1 at 30 SEK/month.

The three most preferred treatment concepts as can be seen in Table 6 are H alternatives, X1 at 30 SEK/month and X1 at 325 SEK/month. The first two of these alternatives are also the most preferred concepts of all the ranked cards, even if we have chosen only to present the ranking between the realistic H and warfarin alternatives.

4.1.3 Test for starting point bias in the conjoint analysis

Conjoint studies tend to be affected by starting point bias, implying that respondents make estimates by starting from an initial value, which is adjusted to yield the final answer. This will lead to valuations that are biased towards the initial value (se Boyle et al. 1985; and Stålhammar, 1996).

To test for starting point bias we asked the patients to answer a question according to the contingent valuation method before they performed the ranking of the conjoint cards. The contingent valuation question is described in Appendix Figure A1. However, with the contingent valuation question we are only testing for anchoring against the bid in the contingent valuation question. An anchoring problem between the different conjoint questions can still exist, and this is not tested for in the analysis.

From Appendix Table A2 and A3 we see that the conjoint ranking is hardly affected at all when we group the patients according to starting bids in the contingent valuation question. In Table A2 the importance levels for the attributes are very similar across the starting bids and in Table A3 the H-like alternatives at 30, 325 and 650 SEK/month are the most prefered treatment concepts for all starting bids, which means that the anchoring effect towards the contingent valuation bids is very small in the conjoint analysis.

4.1.4 Mean willingness to pay by conjoint analysis

The mean willingness to pay for the different attribute levels, as shown in Table 7a and 7b, is easily calculated from the utility function (see Table 2) by dividing the estimated coefficient of a selected attribute level by the negative average marginal utility of money, (see Lancaster, 1971). Mean willingness to pay for a special treatment concept is calculated by taking the sum of the individual mean willingness to pay for the included attribute levels. The average marginal utility of money is -0.0052 according to the average utility function, and using this "price coefficient" we can calculate the mean willingness to pay for each single movement from one attribute level to another within each attribute.

Since the distribution of the mean willingness to pay is unknown in a conjoint analysis, the confidence interval is not directly available when using OLS to estimate the aggregate utility function. However, the confidence interval for the mean willingness to pay has been calculated by means of bootstrap distributions of 1000 so-called pseudo-samples (re-sampling with replacements from the original data set, see Efron and Tisbshirani, 1993; Tambour and Zethraeus, 1998).

The three warfarin treatment concepts W1, W2 and W3 are compared with the H alternative X1 (warfarin alternative W3 is also compared with H alternative X2 for comparison to the contingent valuation offer with "Many limitations in daily life" see Table 10 below).

Table 7a. Mean willingness to pay for movements in attribute levels. 95% confidence interval is shown within brackets.

Changes in attribute levels	**Mean willingness to pay, SEK/month**			
	WTP/ attribute level	**X1 vs. W1**	**X1 vs. W2**	**X1 vs. W3**
C_1 to C_2	162.0			
C_1 to C_3	127.3	127.3		127.3
C_2 to C_3	-34.7		-34.7	
S_1 to S_2	555.4	555.4	555.4	555.4
D_1 to D_2	14.3	-14.3	-14.3	-14.3
L_1 to L_2	299.0			
L_1 to L_3	418.5	418.5		
L_2 to L_3	119.6		119.6	119.6
Mean willingness to pay		**1089[954:1230]**	**626[528:728]**	**788[648:876]**

Table 7b. Mean willingness to pay for movements in attribute levels. 95% confidence interval is shown within brackets.

Changes in attribute levels	**Mean willingness to pay, SEK/month**			
	WTP/ attribute level	**X2 vs. W1**	**X2 vs. W2**	**X2 vs. W3**
C_1 to C_2	162.0			
C_1 to C_3	127.3	127.3		127.3
C_2 to C_3	-34.7		-34.7	
S_1 to S_2	555.4	555.4	555.4	555.4
D_1 to D_2	-14.3	-14.3	-14.3	-14.3
L_1 to L_2	299.0		-299	-299
L_1 to L_3	418.5			
L_2 to L_3	119.6			
Mean willingness to pay		**668[575:764]**	**207[138:285]**	**369[293:453]**

4.2 Contingent valuation

To estimate the mean willingness to pay for an alternative, H-like, treatment as described in Figure A1, we have chosen to use one parametric estimation alternative, logistic regression, and one non-parametric alternative, Kriström's model. The result for the logistic regression is shown in Table 8 below. Illness1 and illness2 are dummy variables for degree of illness where; Illness1 is coded as 1 for patients who got warfarin for a primarily preventive reason and Illness2 is coded as 1 for patients who received warfarin due to the fact that they had had stroke (base line is patients who received warfarin due to the fact that they had had periphery emboli). Gender is coded as 1 for female patients. Age is continuous. Income is a dummy variable coded as 1 for patients who declared that they had a good income (base line is patients who declared that their income was not so good). Work is a dummy variable coded as 1 for patients who declared that they were in gainful employment. Average time for checks is continuous (our hyposesis for this variable is a positive sign, indicating that that patients who spend much time for an average check are more willing to pay for an alternative treatment). Scope variable is a dummy variable coded as 1 for patients who received "Few limitations in daily life" as contingent valuation offer.

Table 8. Logistic regressions, p-values within parenthesis.

Variables	(1) All patients included. Scope variable "Many limitations in daily life" is coded as 0.	(2) Patients included who received the offer: "Many limitations in daily life"	(3) Patients included who received the offer: "Few limitations in daily life"
Constant	0.325 (0.77)	1.365 (0.40)	-0.695 (0.67)
Bid	-0.00190 (0.00)	-0.00199 (0.00)	-0.00179 (0.00)
Illness1	0.0771 (0.7)	0.661 (0.45)	1.005 (0.25)
Illness2	0.923 (0.16)	1.004 (0.28	0.968 (0.30)
Gender	0.241 (0.22)	0.0969 (0.74)	0.387 (0.15)
Age	-0.0169 (0.19)	-0.0297 (0.11)	-0.00141 (0.94)
Income proxy	0.554 (0.003)	0.397 (0.15)	0.670 (0.01)
Work	-0.160 (0.61)	-0.378 (0.46)	0.0621 (0.88)
Average time for checks	0.00339 (0.12)	0.00644 (0.16)	0.00172 (0.52)
Scope variable	0.435 (0.063)		
n	646	329	349
Nagelkerke R2	0.20	0.22	0.19
Log likelihood	-372.10	-172.50	-197.68
Correct prediction	70.90	75.24	66.47
χ^2	105,01	53.18	51.43
E(WTP), SEK/month	**580**	**501**	**661**

Including only significant parameters, at the 5%-level, in the equations does not significantly change resulting mean willingness to pay: 593 SEK/month in equation (1), 523 SEK/month in equation (2) and 664 SEK/month in equation (3).

If the scope effect is tested by setting the scope variable to 0 and 1 respectively when calculating mean willingness to pay in equation (1) we get: 504 SEK/month for scope variable = 0 and 656 SEK/month when scope variable = 1. This is not significantly different from mean willingness to pay in equation (2) and (3) respectively.

From Table 8, column (1) it can be seen that two variables significantly affect the probability of accepting; the Bid and Income.

The estimated parameter for the Scope variable is positive in Table 8, and furthermore, it can be seen that mean willingness to pay in the group who received the offer "Few limitations in daily life", is higher than mean willingness to pay for the group who received the offer "many limitations in daily life". This indicates consistency in the contingent valuation study and is in line with economic theory since you would expect higher mean willingness to pay if the treatment few affects your daily life.

In Kriström's model the mean willingness to pay is established by studying the number of accepts. As we can see in Table 9 still 21% percent of the patients who received the highest bid, 1300 SEK/month, accepted the offer.

Table 9. Willingness to pay estimated by Kriströms (2) method, All patients.

Bid	Interval	Midpoint	Number of accepts in the offer	Total	Fraction of accepts	E(WTP)
	0	0	-	-	1	0
100	0-100	50	119	168	0.71	14.58
500	100-500	300	51	166	0.31	120.33
900	500-900	700	45	171	0.26	30.85
1300	900-1300	1100	36	171	0.21	57.79
	Extrapolation: 1300-2980	2140	0	0	0	450.52
Total			251	676		
E(WTP)						**674**

It can also be seen from Table 9 that between the offers 500 to 900 SEK/month, the number of accepters has decreased by 5%, and this is also the case between the offers 900 to 1300 SEK/month. If this 'trend' is used for extrapolation, we will have no further accepts when the price is 2980 SEK/month, and the mean willingness to pay will be 674 SEK/month (called Kriström (2) in Table 10 below). If the highest bid, 1300SEK/month, is used for truncation, the mean willingness to pay will be 539 SEK/month (called Kriström (1) in Table 10.

4.3 Conjoint analysis versus contingent valuation

To test how the scope variable "Limitations in daily life" affects the willingness to pay in the contingent valuation study the patients have been divided into two groups: patients who received the offer "Few limitations in daily life" in the contingent valuation question and patients who received the offer "Many limitations in daily life" in the contingent valuation question.

Confidence interval for mean willingness to pay estimated by logistic regression and Kriström's method are calculated by means of bootstrapping in the same manner as described in the conjoint study above (see Efron and Tisbshirani, 1993).

With contingent valuation-offer "Many limitations in daily life", we can see from Table 10 below that mean willingness to pay lies in the interval 369 to 603 SEK/month. Mean willingness to pay estimated by Kriström (2) in the contingent valuation analysis and by logistic regression are significantly different at the 5%-level from mean willingness to pay estimated by conjoint analysis when tested by a t-test.

Expected mean willingness to pay lies in the interval 593 to 788 SEK/month for patients with contingent valuation-offer "Few limitations in daily life" if we take the W3 treatment concept as the most realistic warfarin-like treatment today compared to X1 as the most realistic H-like alternative (see for example Eriksson, 2002; Olsson, 2003; and Wallentin et al. 2003). Then the mean willingness to pay estimated by the conjoint method compared to mean willingness to pay measured by all three methods in the contingent valuation study are not significantly different from each other at the 5%-level when tested by a t-test.

As we can see from Table 10 above, the difference in the scope effect is significantly larger at all reasonable levels when tested by a t-test for conjoint analysis compared to all three models used in the contingent valuation analysis.

Table 10. Summary: estimated mean willingness to pay in the study. 95% confidence interval is shown within brackets. p-values are shown on t-test of significant differences between mean willingness to pay estimated by respective model compared to mean willingness to pay estimated with conjoint analysis.

Estimation methods	**WTP estimates and t-tests of differences between the models:** p-values are shown on t-test of significant differences between mean willingness to pay estimated by respective model compared to mean willingness to pay estimated with conjoint analysis					
	offer: "Many limitations in daily life"		offer: "Few limitations in daily life"		Scope effect	
	Mean WTP	p- values on t-test	Mean WTP	p- values on t-test	Estimated scope effect	p- values on t-test
Kriström (1)	**484** [394:609]	**0,23**	**593** [500:712]	**0,08**	**109**	**0,00**
Kriström (2)	**603** [465:796]	**0,00**	**744** [602:930]	**0,51**	**141**	**0,00**
Logistic Regression	**501** [462:597]	**0,04**	**661** [612:745]	**0,14**	**159**	**0,00**
Conjoint	**X2 vs. W3** **369** [293:453]		**X1 vs. W3** **788** [648:876]		**419**	

5 Conclusions

Cardio-vascular diseases cause 50% of all deaths in Sweden, and the most common oral anti coagulant (OA) treatment today is warfarin treatment, giving protection against severe illness but involving a high risk of side effects. A new drug, H, has proven in clinical tests to be an effective drug involving fewer INR checks, the same dose all the time, fewer side effects and fewer limitations in daily life for the patients (see Eriksson, 2002; Olsson, 2003; and Wallentin et al. 2003).

This study aims firstly to analyze preferences for different AC-treatments among 682 patients treated with warfarin for at least 6 months. For this purpose the conjoint approach has been frequently used (see for example Ryan and Hughes, 1997; Johnson et al. 1998; Aristides et al. 2004; and Telser and Zweifel, 2002). Results show that the treatment characteristics preferred in the alternative OA treatment were fewer side effects and fewer limitations in daily life. Characteristics such as number of checks and dose of the drug were of less importance. Furthermore, two H-like treatment concepts, at 30 SEK/month and 325 SEK/month, receive the highest utility.

Secondly, in the decision as to whether H should be included in health insurance, mean willingness to pay estimates are of great importance, and the case for either the conjoint approach and the contingent valuation approach can be proposed (see Telser and Zweifel, 2002). In this study we actually compare the two methods to each other.

There are not many studies to be found on the comparison of the contingent valuation and the conjoint techniques. Other studies to be mentioned are for example Stevens et al. (2000) who studied willingness to pay for ecosystem management, the estimation of willingness to pay for water purification programs of Barrett et al. (1996), the study of Boxall et al. (1996) on willingness to pay for moose hunting, the willingness to pay study for risk reduction of Magat et al. (1988) and the Desvouges and Smith (1983) study on water quality. In all the above studies the estimation of willingness to pay by the two techniques differed substantially. Boxall et al. showed willingness to pay estimated by contingent valuation higher than those estimated by the conjoint method whereas the other studies obtained the opposite result.

According to Stevens et al. (2000) conjoint willingness to pay estimates might be biased upwards since conjoint studies count "maybe" responses as "yes" responses, while Magat et al. (1988) argue that the willingness to pay estimates by the contingent valuation approach would be biased downwards since this technique does not provide the respondent with a trade-off situation.

In the contingent valuation question in the present study, patients were randomly given the offers "Many limitations in daily life" and "Few limitations in daily life". For patients who were given the contingent valuation-offer "Many limitations in daily life" mean willingness to pay lies in the interval from 484 to 603 SEK/month estimated by logistic regression and Kriström's method. The corresponding estimate with conjoint analysis is 369 SEK/month.

Furthermore, mean willingness to pay lies in the interval from 593 to 744 SEK/month for patients who were given the contingent valuation-offer "Few limitations in daily life" with logistic regression and Kriström's method. With conjoint analysis mean willingness to pay becomes 788 SEK/month.

Also woth noticing is that the difference in the scope effect is significantly larger when we use conjoint analysis compared to the methods used in contingent valuation analysis.

This paper indicates that there is a substantial willingness to pay for a H-like treatment concept with conjoint analysis as well as with contingent valuation analysis, suggesting that H could be included in the reimbursement system. However, our study differs compared to the other studies mentioned above in that mean willingness to pay estimated by the conjoint method and the contingent valuation method are close.

References

Aristides, M., A. R. Weston, P. FitzGerald, C. Le Reun and N.Maniadakis (2004). Patient Preference and Willingness-to-Pay for Humalog Mix25 Relative to Humulin 30/70: A Multicountry Application of a Discrete Choice Experiment. Value in Health, 7, (4), 442–454.

Asplund, K. (1995). Förmaksflimmer: Epidemiologi, Morbiditet och Prognos. Förmaksflimmer, *Klinisk rutin*.

Boyle, K. J., R. C. Bishop and M. P. Welsh (1985). Starting Point Bias in Contingent Valuation Bidding Games. *Land Economics*, 61, 188-194.

Cattin, P. and D. Wittink (1982). Commercial use of conjoint analysis: a survey. *Journal of Marketing*, 46, 44–53.

Coursey, D. L. and W. D. Schulze (1987). The Disparity Between Willingness to Accept and Willingness to Pay Meassures of Value. *The Quarterly Journal of Economics*, 102, 679-690

Efron, B. and R. J. Tisbshirani (1993). *An Introduction to the Bootstrap. Monographs on Statistics and Applied Probability*. New York, London: Chapman & Hall. No 57.

Eriksson, H. (2002). Oral presentation of THRIVE III in Philadelphia, American Society of Hematology.

Eriksson, S., K. Asplund, E. Hägg, F. Lithner, T. Strand and P. O. Wester (1987). *Clinical profiles of cerebrovascular disorder in a populationbased patient sample*, J Chron Dis, 40, 1025-1032.

Green, P. E. and V. Srinivasan (1990). Conjoint analysis in marketing: New developments with implications for research and practice. *Journal of Marketing*, 54(4), 3-19.

Gustafsson, A., A. Herrmann and F. Huber (2001). *Conjoint Measurement, Methods and Applications*, Second Edition, Springer.

Hanemann, W. M. (1994). Valuing the environment through contingent valuation. *Journal of Economic Perspectives*, 8, 19-43.

Herriges, J. A. and J. F. Shogren (1996). Starting Point Bias in Dichotomous Choice Valuation with Follow-up Questioning. *Journal of Environmental Economics and Management*, 30, 112-131.

Jain, A. R., F. Acito, N. Malhorta and V. Mahajand (1979). "A Comparison of Internal Validity of Alternative Paremater Estimation Methods in Decompositional Multiattribute Preference Models", *Journal of Marketing Research*, 16, 313-322.

Johansson, G., B. Stålberg, G. Törnling, S. Andersson et al. (2004). Asthma treatment preference study: a conjoint analysis of preferred drug treatments - clinical investigations. *Chest*, March.

Johansson, P. O. (1995). *Evaluating Health risk, An Economic Approach*. Cambridge, Cambridge University Press.

Johnson, F. R., M. C. Ruby, M. H. Desvousges and J. R. King (1998). Using stated preferences and health-state classifications to estimate the value of health effects of air pollution. TER Working Paper, No. T-9807.

Johnsson, H. (1999). Antikoagulantiabehandlingen kan göras säkrare, *Läkartidningen*, 96, 32-33.

Johannesson, M. (1993). The contingent valuation method: appraising the appraisers. *Health Economics*, 2, 357-359.

Johannesson, M. (1996). The contingent valuation controversy in environmental economics and its relevance to health services research. *Journal of Health Services Research & Policy*, 1, 116-117.

Johannesson, M. and B. Jönsson (1991). Economic evaluation in health care: is there a role for cost-benefit analysis? *Health Policy*, 17, 1-23.

Kendall, M. G. (1970). *Rank Correlation Methods*. London, Griffin.

Kriström, B. (1990). A Non-Parametric Approach to the Estimation of Welfare Measures in Discrete Response Valuation Studies. *Land Economic*, 66, (2), 135-139.

Lancaster, K. (1971). *Consumer Demand: a New Approach*. New York, Columbia University Press.

McFadden, D. (1994). Contingent valuation and social choice. *American Journal of Agricultural Economics*, 76, 689-708.

National Oceanic and Atmospheric Administration. (1994). Report of the NOAA Panel on Contingent Valuation. *Federal Register*, 58, 4601-4614.

Olsson, S. B. (2003). Oral presentation of SORTIF III in Vienna, European Society Cardiology.

Osman, L. M., L. McKenzie, J. Cairns et al. (2001). Patient Weighting of Importance of Asthma Symptoms. *Thorax*, 56,138-142.

Petersen, P. (1990). Thromboembolic complications in atrial fibrillation. *Stroke*, 21, 4-13.

Ryan, M. (1996). *Using consumer preferences in health care decision making: the application of conjoint analysis*. Office of Health Economics. Luton, Bedfordshire, UK: White Crescent Press.

Ryan, M. and J. Hughes (1997). Using Conjoint Analysis to Assess Women's Preferences for Miscarriage Management. Health Economics, 6, 261–273.

Själander, A., G. Engström, E. Berntorp and P. Svensson (2003). Risk of haemorrhagic stroke in patients with oral anticoagulation compared to the general population. *Journal of International Medicine*, Nov, 254 (5), 434-438.

SPSS, 2004. Advanced Market Research.

Stålhammar, N. O. (1996). An Empirical Note on Willingness to Pay and Starting Point Bias. *Medical Decision Making*, 16, 242-247.

Szeinbach, S. L., J. H. Barnes, W. F. McGhan et al. (1999). Using conjoint analysis to evaluate health state preferences. *Drug Inf Journal*, 33, 849-858.

Tambour, M. and N. Zethraeus (1998). Nonparametric Willingness-to-pay Measures and Confidence Statements. Medical Decision Making, 18, 330-336.

Terent, A. (1998). Dyrt med felinställd antikoagulantiaterapi. Kan egenkontroll förbättra situationen?. *Läkartidningen*, 95, 2233-2238.

Telser, H. and P. Zweifel (2002). Meassuring willingness-to-pay for risk reduction: an application of conjoint analysis. *Health Economics,* 11, 129-139.

Thayer, M. A. (1981). Contingent Valuation techniques for Assessing Environmental Impacts: Further Evidence. *Journal of Environmental Economics and Management*, 8, 27-44.

Wallsten, T. (1967). Using Conjoint-Measurement models to investigate a theory about probabilistic information processing. *Journal of Mathematical Psychology*, 14, 144-185.

Wallentin, L., R. G.Wilcox, W. D. Weaver et al. (2003). Oral H for secondary prophylaxis after myocardial infarction: the ESTEEM randomised controlled trail. *Lancet* 362, 789-97.

Appendix

Table A1. Questionnaire on patient variables.

Questions answered by nurses or physicians:	**Answers:**
Date of today's warfarin check	YYYY MM DD
Year of birth	YYYY MM DD
Gender	Female/Male
Is the patient taking warfarin for primarily preventive reason, i.e. the patient has not had stroke or periphery emboli?	Yes/No
If no, why is the patient taking warfarin?	Stroke/periphery emboli
Date for prescription of warfarin treatment?	YYYY MM DD
How many warfarin checks has the patient had during the last three months, today's check included?	
Questions answered by the patients:	**Answers:**
How did you travel there?	Taxi (disabeled transport)/ Public transport/ Car (own or relatives)/ Other way
Have you taken time off work to attend this appointment?	Yes/No/I don't work
Has anyone helped you to get there?	A relative/Social worker/No
If a relative or friend helped you, did this person take time off work to do so?	Yes/No
What is your occupation?	Studying/ Full time employment or own company/ Part time employment or own company/ Retirement due to illness/ Retirement due to age/ Long time illness/ Unemployed/ Other
On average, how long does it take you to attend a warfarin check (total time inclusive travelling?	
How would you grade your private income?	Very good/ Relatively good/ Neither good nor bad/ Relatively bad/ Very bad

Table A2. Study of starting point bias in the conjoint analyses. Importance (in percentage) for the attributes for patients grouped according to bids in the contingent valuation question

	Patients grouped by bids, in SEK/month, in the offer Figures are importance in percentage			
	100	**500**	**900**	**1300**
Monthly cost	43.8	45.3	46.2	40.4
Side effect	26.7	25.3	25.8	28.2
Limitations in daily life	21.1	19.4	18.8	21.5
Checks	7.3	8	8.4	8.6
Dose	1.1	2	0.8	1.3

Table A3. Study of starting point bias in the conjoint analyses. Most preferred treatment concepts in different groups based on bids in the contingent valuation question. Monthly cost in SEK in parenthesis.

Different bids, in SEK/month, in the offer			
100	**500**	**900**	**1300**
X1 (30)	X1 (30)	X1 (30)	X1 (30)
X1 (325)	X1 (325)	X1 (325)	X1 (325)
X1 (650)	X1 (650)	X1 (650)	X1 (650)
W2 (30)	W2 (30)	W2 (30)	W2 (30)
W3 (30)	W3 (30)	W3 (30)	W3 (30)
W2 (325)	X1 (975)	W2 (325)	W2 (325)
X1 (975)	W2 (325)	W3 (325)	X1 (975)
W3 (325)	W3 (325)	X1 (975)	W3 (325)

Figure A1. The contingentvaluation offer.

You have atrial fibrillation, which involves a higher risk of thromboembolism that can cause stroke. To prevent this risk you are treated with warfarin - an anti coagulant drug.

Assume that there is an alternative treatment to your current one. This new treatment alternative gives the same protection against thrombi embolism and stroke as your current treatment. This means that even if you perceive improvements in some aspects on this hypothetical treatment alternative compared to your current treatment, you will not run a higher risk of thromboembolism or stroke.

Assume that the drug that protects you from thrombi embolism and stroke involves the following:

Number of checks
One blood test every month the first six months, after that no more checks.

Side effects
Your drug is unlikely to interact with other drugs, which lesses the risk of minor haemorrhage such as nose bleed, bleeding gums and bruising.

Dose
One pill twice a day, and the number of pills never changes.

Limitations in daily life
The following attribute levels were randomised among patients: Many
Few

We are interested in your opinion. This will not affect the cost of your medication in the future.
Assume that the treatment presented above would cost you (the following prices were randomized among the patients): 100, 500, 900, 1300 SEK per month out of your own pocket. Would you be willing to pay this extra cost to get the proposed treatment instead of your current treatment?

☐ **Yes**

☐ **No**

You may give reasons for your answer here:

EFI, The Economic Research Institute

Published in the language indicated by the title. A complete publication list can be found at www.hhs.se/efi
Books and dissertations can be ordered from EFI via e-mail: EFI.Publications@hhs.se

Reports since 2001

2006

Books

Wijkström, Filip och Torbjörn Einarsson. *Från nationalstat till näringsliv? Det civila samhällets organisationsliv i förändring.*

Dissertations

Gaspar, Raquel M. *Credit Risk and Forward Price Models.*
Hopkins, Elisabeth. *Is a Higher Degree of Local Currency Pricing Associated With Lower Exchange Rate Pass-Through? A Study of Import Pricing in 51 Swedish Industries.*
Kling, Ragnar. *Developing Product Development in Times of Brutal Change.*
Langenskiöld, Sophie. *Peer Influence on Smoking: Causation or Correlation?*
Meitz, Mika. *Five Contributions to Econometric Theory and the Econometrics of Ultra-High-Frequency Data.*
Wilander, Fredrik. *Essays on Exchange Rates and Prices.*

2005

Books

Andersson, Per, Susanne Hertz and Susanne Sweet (eds). *Perspectives on market networks – boundaries and new connections.*
Charpentier, Claes. *IT inom omsorgen. Förväntade effekter av införande av IT-system för utförarna inom äldre- och handikappomsorgen.*
Dembrower, Maria. *Entreprenörskap i industriella nätverk.*
Lind, Johnny och Göran Nilsson (redaktörer). *Ekonomistyrningens metoder, sammanhang och utveckling – En vänbok till Lars A Samuelson.*
Samuelson, Lars A. *Organizational governance and control – a summary of research in the Swedish society.*

Dissertations

Andersson, Martin. *Making a Difference – Project Result Improvement in Organizations.*
Arvidsson, Per. *Styrning med belöningssystem – Två fallstudier om effekter av belöningssystem som styrmedel.*
Bems, Rudolfs. *Essays in International Macroeconomics.*
Berg-Suurwee, Ulrika. *Nya trender, nya nämnder – effekter av en stadsdelsnämndsreform inom kultur och fritid.*
Björkman, Hans. *Learning from members – Tools for strategic positioning and service innovation in trade unions.*
Bodnaruk, Andriy. *Essays in Empirical Corporate Finance and Portfolio Choice.*
Clapham, Eric. *Essays in Real Estate Finance.*
Dareblom, Jeanette. *Prat, politik och praktik – Om individers möten med strukturer i en kommunal satsning på kvinnors företagande.*

Ögren, Anders. *Empirical Studies in Money, Credit and Banking – The Swedish Credit Market in Transition under the Silver and the Gold Standards, 1834–1913.*

2002

Books

Schuster, Walter. *Företagets Valutarisk – En studie av horisontella och vertikala styrprocesser.*

Sjöstrand, Sven-Erik och Pernilla Petrelius. *Rekrytering av koncernstyrelsen – Nomineringsförfaranden och styrelsesammansättning med fokus på kvinnors ställning och möjligheter.* EFI/SNS Förlag

Löwstedt, Jan och Bengt Stymne (red). *Scener ur ett företag – Organiseringsteori för kunskapssamhället.* EFIs Årsbok 2002.EFI/Studentlitteratur.

Dissertations

Barinaga, Ester. *Levelling Vagueness – A Study of Cultural Diversity in an International Project Group.*

Berglund, Johan. *De otillräckliga – En studie av personalspecialisternas kamp för erkännande och status.*

Bolander, Pernilla. *Anställningsbilder och rekryteringsbeslut.*

Damjanovic, Tatiana. *Essays in Public Finance.*

Ekman, Mattias. *Studies in Health Economics – Modelling and Data Analysis of Costs and Survival.*

Heyman, Fredrik. *Empirical Studies on Wages, Firm Performance and Job Turnover.*

Kallifatides, Markus. *Modern företagsledning och omoderna företagsledare.*

Kaplan, Michael. *Acquisition of Electronic Commerce Capability – The Cases of Compaq and Dell in Sweden.*

Mähring, Magnus. *IT Project Governance.*

Nilsson, Mattias. *Essays in Empirical Corporate Finance and Governance.*

Schenkel, Andrew. *Communities of Practice or Communities of Discipline – Managing Deviations at the Øresund Bridge.*

Skogsvik, Stina. *Redovisningsmått, värderelevans och informationseffektivitet.*

Sundén, David. *The Dynamics of Pension Reform.*

Ternström, Ingela. *The Management of Common-Pool Resources – Theoretical Essays and Empirical Evidence.*

Tullberg, Jan. *Reciprocitet – Etiska normer och praktiskt samarbete.*

Westling, Gunnar. *Balancing Innovation and Control – The Role of Face-to-face Meetings in Complex Product Development Projects.*

Viklund, Mattias. *Risk Policy – Trust, Risk Perception, and Attitudes.*

Vlachos, Jonas. *Risk Matters – Studies in Finance, Trade and Politics.*

Mattsson, Susanna. *På gränsen mellan ordning och oordning – tingens betydelse vid marknadsombildningar. En studie av svenska postväsendets ombildning under 1990-talet.*
Nilsson, Charlotte. *Studies in Environmental Economics: Numerical Analysis of Greenhouse Gas Policies.*
Nilsson, Hans. *Medborgaren i styrsystemet – beskrivning av VAD och HUR i styrning av kommunal verksamhet.*
Nystedt, Jens. *Competition, Regulation and Integration in International Financial Markets.*
Pajuste, Anete. *Corporate Governance and Controlling Shareholders.*
Richtnér, Anders. *Balancing Knowledge Creation. Organizational Slack and Knowledge Creation in Product Development.*
Salabasis, Mickael. *Bayesian Time Series and Panel Models – Unit Roots, Dynamics and Random Effects.*
Sandberg, Rickard. *Testing the Unit Root Hypothesis in Nonlinear Time Series and Panel Models.*
Skallsjö, Sven. *Essays on Term Structure and Monetary Policy.*
Strikholm, Birgit. *Essays on Nonlinear Time Series Modelling and Hypothesis Testing.*
Söderström, John. *Från Produkt till Tjänst. Utveckling av affärs- och miljöstrategier i produktorienterade företag.*
Talia, Krim. *The Scandinavian Currency Union, 1873–1924 – Studies in Monetary Integration and Disintegration.*

2003

Books

Lundahl, Mats (editor). *Globalization and Its Enemies.* EFIs Årsbok 2003.
Sundgren, Bo, Pär Mårtensson, Magnus Mähring and Kristina Nilsson (editors). *Exploring Patterns in Information Management. Concepts and Perspectives for Understanding IT-Related Change.*

Dissertations

Andersson, Henrik. *Valuation and Hedging of Long-Term Asset-Linked Contracts.*
Bergman, Malin. *Essays on Human Capital and Wage Formation.*
Damsgaard, Niclas. *Deregulation and Regulation of Electricity Markets.*
Eklund, Bruno. *Four Contributions to Statistical Inference in Econometrics.*
Hakkala, Katariina. *Essays on Restructuring and Production Decisions in Multi-Plant Firms.*
Holgersson, Charlotte. *Rekrytering av företagsledare. En studie i homosocialitet.*
Ivaschenko, Iryna. *Essays on Corporate Risk, U.S. Business Cycles, International Spillovers of Stock Returns, and Dual Listing.*
Lange, Fredrik. *Brand Choice in Goal-derived Categories – What are the Determinants?*
Le Coq, Chloé. *Quantity Choices and Market Power in Electricity Market.*
Magnusson, Peter R. *Customer-Oriented Product Development – Experiments Involving Users in Service Innovation.*
Meisiek, Stefan. *Beyond the Emotional Work Event Social Sharing of Emotion in Organizations.*
Mårtensson, Anders. *Managing Mission-Critical IT in the Financial Industry.*
Nilsson, Göran. *Processorientering och styrning – Regler, mål eller värderingar?*
Sandberg, Robert. *Corporate Consulting for Customer Solutions Bridging Diverging Business Logics.*
Sturluson, Jon Thor. *Topics in the Industrial Organization of Electricity Markets.*
Tillberg, Ulrika. *Ledarskap och samarbete – En jämförande fallstudie i tre skolor.*
Waldenström, Daniel. *Essays in Historical Finance.*
Wallén, Ulrika. *Effektivitet i grundskolan i anslutning till en stadsdelsnämndsreform.*

Fromm, Jana. *Risk Denial and Neglect: Studies in Risk Perception.*
Hjelström, Anja. *Understanding International Accounting Standard Setting – A Case Study of IAS 12, Accounting for Deferred Tax.*
Hortlund, Per. *Studies on Swedish Banking 1870-2001.*
Langenskiöld, Sophie. *Peer Influence on Smoking: Causation or Correlation?*
Lindahl, Therese. *Strategic and Environmental Uncertainty in Social Dilemmas.*
Linnarsson, Håkan. *Alliance for Innovation. A structural perspective on new business development in cooperative ventures.*
Madestam, Andreas. *Developing Credit Markets*
Nilsson, Roland. *The Market Impact of Short-Sale Constraints.*
Nordfält, Jens. *Is consumer decision-making out of control? Non-conscious influences on the consumer decision-making process for fast moving consumer goods.*
Nordin, Fredrik. *Externalising Services – Walking a Tightrope between Industrial and Service Logics.*
Parmler, Johan. *Essays in Empirical Asset Pricing.*
Simbanegavi, Witness. *Price Discrimination, Advertising and Competition.*
Thodenius, Björn. *Användning av ledningsinformationssystem: en longitudinell studie av svenska storföretag.*
Tolis, Christofer. *Framing the Business – Business Modelling for Business Development.*
Östberg, Per. *Corporate Disclosure and Investor Recognition.*

2004

Books

Ahrne, Göran och Nils Brunsson (red). *Regelexplosionen.*
Lind, Johnny. *Strategi och ekonomistyrning. En studie av sambanden mellan koncernstrategi, affärsstrategi och ekonomistyrning.*
Lind, Johnny och Walter Schuster (red). *Redovisningens teori, praktik och pedagogik. En vänbok till Lars Östman.*
Sevón, Guje och Lennart Sjöberg (red). *Emotioner och värderingar i näringslivet.* EFIs Årsbok 2004.
Wijkström, Filip and Stefan Einarsson. *Foundations in Sweden – Their scope, roles and visions.*

Dissertations

Anderson, Anders. *Essays in Behavioral Finance.*
Balsvik, Gudrun. *Information Technology Users: Studies of Self-Efficacy and Creativity among Swedish Newspaper Journalists.*
Blix, Magnus. *Essays in Mathematical Finance – Modelling the Futures Price.*
González Gómez, Andrés. *Nonlinear dynamics and smooth transition models.*
Grönqvist, Erik. *Selection and Moral Hazard in Health Insurance: Taking Contract Theory to the Data.*
Ivarsson Westerberg, Anders. *Papperspolisen – varför ökar administrationen i moderna organisationer.*
Jutterström, Mats. *Att påverka beslut – företag i EUs regelsättande.*
Jönsson, Kristian. *Macroeconomic Aspects of Capital Flows to Small Open Economies in Transition.*
Larsson, Pär. *Förändringens villkor. En studie av organisatoriskt lärande och förändring inom skolan.*
Lagerwall, Björn. *Empirical Studies of Portfolio Choice and Asset Prices.*
Malmsten, Hans. *Properties and Evaluation of Volatility Models.*
Marshall, Cassandra. *Dating for Innovation. Creating and Recognizing Opportunities through Collaborative Interorganizational Relationships in Fluid Environments.*