ISSN 1503-299X

WORKING PAPER SERIES

No. 25/2002

NON-LINEAR UNEMPLOYMENT EFFECTS IN SICKNESS ABSENCE: DISCIPLINE OR COMPOSITION EFFECTS?

Jan Morten Dyrstad
Solveig Osborg Ose

Department of Economics

Norwegian University of Science and Technology
N-7491 Trondheim, Norway
www.svt.ntnu.no/iso/wp/wp.htm

Non-linear Unemployment Effects in Sickness Absence: Discipline or Composition Effects?*

Jan Morten Dyrstad   Solveig Osborg Ose

Department of Economics

Norwegian University of Science and Technology

N-7491 Trondheim, Norway

February 2001

Abstract

By applying Smooth Transition Regressions (Teräsvirta, 1998) we test whether effects on short (STA) and long term (LTA) sickness absence depend on the *level* of unemployment. The main question is to what extent unemployment affects sickness absence through so-called discipline and/or composition effects. The empirical analysis is carried out on time series data referring to blue collar workers in Norway. For LTA we conclude that only discipline effects are present. Concerning STA, none of these effects seems present among women, whereas both effects may be present among men. Non-linear effects of wage and sick pay scheme changes support these conclusions.

*Comments from seminar participants at Department of Economics, NTNU, and Stavern June 2000 are gratefully acknowledged. We are grateful to Confederation of Norwegian Business and Industry (NHO) for financial support.

JEL Classification: C22, J22, J33, J64

Keywords: sickness absence, unemployment, non-linear modelling

1 Introduction

A number of studies, based on data from different countries, report that increased unemployment reduce sickness absence.[1] At least from the viewpoint of economics, discipline and composition mechanisms are the two main explanations of a negative relationship between sickness absence and unemployment.[2]

Unemployment may operate as a *discipline* mechanism because workers in periods of high unemployment are more likely to go to work even if they are sick. The reason is that workers with "bad" absence records are more likely to be dismissed than those with "good" records, and it is more difficult to find a new job when unemployment is high. Furthermore, in situations of high unemployment employers have more people to choose among, so it becomes less risky to fire workers because it is easier to find new workers.

The *composition* mechanism can be explained as follows: Under the assumption that labour supply is constant, increased unemployment means that firms go out of business but also that firms carry out manning reductions. If firms dismiss so-called "marginal workers", *e.g.*, workers with health problems, the average health condition among those at work improves so that sickness absence is reduced.[3]

The main aim of this paper is to identify these two effects empirically by investigating whether changes in unemployment affect sickness absence differently in periods of "high" unemployment as compared to periods of "low" unemployment.

It is both interesting and important to separate these two effects empirically. High rates of sickness absence means that a corresponding large part of the working force is

[1] On time series data Doherty (1979), Lantto and Lindblom (1987), Kenyon and Dawkins (1989), and Dyrstad and Lysø (1998) find evidence of a negative relationship between absence and unemployment. Based on cross-section data the studies by Allen (1981), Chaudhury and Ng (1992), Drago and Wooden (1992), and Johanson and Palme (1996) also get results supporting the negative relationship.

[2] A third is that working pressure is reduced when unemployment is high because low unemployment corresponds to low demand in product markets and consequently reduced production. So under the reasonable assumption that reduced working pressure reduces sickness absence, this may also explain a negative relationship between sickness absence and unemployment. On the other hand, it has been mentioned that higher unemployment may *increase* absence, because increased unemployment may create psychical problems among workers due to increased economic and social uncertainty, which lead to more sickness absence.

[3] To this it should be added that it is *formally* illegal in Norway and other countries to dismiss a worker for the reason that she or he has a high rate of sickness absence. However, we think that there are informal channels that can be used to get rid of the least productive workers in case of manning reductions, for instance by changing organisational structure. Refering to Norway, there is some research on demand for disability pensions which by analogy supports this. Dahl (1990), Halvorsen and Johannesen (1991) and Hippe and Pedersen (1991) find that in the end of the 1980s, when unemployment started to rise sharply in Norway, several firms wishing to carry out manning reductions, were able to transfer elderly workers to diability pension schemes.

out of work. It is obviously neither an aim nor possible to bring absence rates down to zero, but absence rate changes will change effective labour supply, and in this respect it is as important as lowering unemployment rates. Furthermore, sickness absence represents costs, to the employees and employers directly involved, and to society as a whole. Therefore concern for sickness absence often appear in public debate, and governments take steps to affect sickness absence. For instance, in the first part of the 1990s the government made several minor changes in the Norwegian sick pay scheme to reduce sickness allowances paid by National Insurance, which in 1991 was still at the peak of 1.8 per cent of gross domestic product. Recent calculations by Hem (2000) indicate that on average a lost working day in Norway due to sickness absence increases firms' cost by 120% beyond wage costs. For firms in manufacturing industries and building and constructions the costs are even higher. For employees in general, sickness absence represents personal suffering and deterioration of skills, thereby reducing competition in the labour market and the possibilities of having higher future income.

In a situation with increasing sickness absence it is therefore often called upon policy measures to bring it down. If increases in sickness absence appear when unemployment is falling, it is important to know what kind of mechanisms are in operation in order to design effective policy measures. If the increases are most likely to be explained by composition mechanisms, one would reasonably think of measures directed towards firms and employees. One example could be incentives stimulating firms to improve working conditions so that "marginal" workers are able to work. Another example could be training programs. In Norway a special programme designed to combine absence and work at the individual level was introduced in 1991. On the other hand, if higher absence is due discipline mechanisms, it seems more reasonable to design macro measures, *e.g.*, making compensation schemes more restrictive.

Our paper does not only focus on effects of changes in unemployment, and the identification of composition and discipline mechanisms. Also possible non-linear effects of changes in the sick pay scheme and wages are addressed.

Empirically we apply the Smooth Transition Regression (STR) approach (Teräsvirta, 1998), both to identify the composition and discipline mechanisms, and to model other non-linearities in sickness absence. An important assumption behind our identification approach is that the discipline effect is stronger if the level of unemployment is "high" as compared to “low” levels of unemployment. This assumption and the hypothesis of composition mechanism identify the discipline and composition effects empirically. The empirical analysis is carried out on quarterly time series data covering the period 1971 (3) - 1998 (2).[4] Data refer to blue collar workers, separated w.r.t. gender, and we use series for both long and short term absence spells.

The plan of the paper is as follows. We use a standard model of individual labour supply to discuss the discipline mechanism, and a discussion of this model and its relation to composition mechanisms is given in Section 2. Section 3 and 4 present data on sickness absence and unemployment, and the empirical analysis, respectively. Estimation results are given and discussed in Section 5. Section 6 contains conclusions.

2 Hypotheses

We may discuss discipline effects within the standard model of individual labour supply,[5] having a representative worker with a strictly quasi-concave utility function with levels of consumption and leisure as arguments, and the following budget restriction:

$$Consumption = W[h - (1-k)A] - P(A, U). \tag{1}$$

The variable A is number of hours absent from work, W is the real disposable wage rate, h is standard working hours and k is degree of economic compensation if absent from work. If k=1 the worker is fully compensated, while k=0 implies no compensation.

The penalty function $P(A, U)$ is a central part of the model as it causes the possible non-linearities we are investigating. We have formulated penalty to depend on individual absence and the aggregate rate of unemployment, U, and we assume that it is convex and increasing in A, *i.e.*, both $\partial P/\partial A$ and $\partial^2 P/\partial A^2$ are positive. Furthermore, we assume that $P(0, U) = 0$.[6] The reason for including penalty in the budget restriction is that absence represents costs to firms. Firms may reduce costs by creating expectations of reduced bonuses, fringe benefits and career opportunities for workers with high rates of

[4] The number in paranthesis refers to quarter.

[5] This is a common theoretical framework in absence research by economists, see Brown and Sessions (1996). However, discipline mechanisms could also be discussed within efficiency wage models, as in Barmby *et al.* (1994).

[6] A unique solution to utility maximization is possible even if $\partial^2 P/\partial A^2 = 0$, but it seems reasonable to assume that it is strict positive.

absence. Increased possibilities of being dismissed if the firm has to reduce employment are particularly important. Penalty in terms of reduced income therefore depends on labor market conditions so that penalty is higher when unemployment is high, also at the margin. Hence, $\partial P/\partial U > 0$ and $\partial^2 P/\partial A \partial U > 0$.

Given the budget restriction (1) and a standard time restriction, utility maximization gives the condition that the marginal rate of substitution between leisure and consumption equals the price of leisure, which is also the price of being absent from work. From the first order conditions the representative worker's absence relation may in general be expressed as

$$A = A(W, k, U, h). \tag{2}$$

Assuming that leisure (absence) and consumption are normal goods, we obtain from this model unambiguous predictions of unemployment changes and degree of economic compensation, *i.e.*, $\partial A/\partial U < 0$ and $\partial A/\partial k > 0$. However, a higher wage rate has an ambiguous effect on absence because the income and substitution effects work in opposite directions. A sufficient condition for a positive relation between absence and standard working hours is that marginal utility of leisure is not reduced if consumption increases.

The central question in this paper concerns the form of the relationship between absence and unemployment in equation (2). Based on the above model, the answer depends on the penalty function. Technically, the sign of $\partial^2 A/\partial U^2$ depends on the signs of $\partial^2 P/\partial U^2$ and two third order derivatives of the penalty function. If we assume that the two third order derivatives are negligible, the relationship between absence and unemployment is concave (convex) if the penalty function is convex (concave) in unemployment.[7]

If the penalty function is convex in unemployment, this means that an increase in unemployment "punish" more when unemployment is "high" as compared to a "low" level of unemployment. The relationship between absence and unemployment is then concave.

[7] The two third order derivatives are $\partial \left(\partial^2 P(\cdot)/\partial A \partial U\right)/\partial U$ and $\partial \left(\partial^2 P(\cdot)/\partial A^2\right)/\partial U$. Setting these equal to zero gives $\partial^2 A(\cdot)/\partial U^2 = (\partial A(\cdot)/\partial Y) \times (\partial^2 P(\cdot)/\partial U^2)$, saying that it depends on the income effect and the form of the relationship between penalty and unemployment. So if these third order derivatives are zero and leisure is a normal good, it follows that the form of the relationship between absence and unemployment depends on $\partial^2 P(\cdot)/\partial U^2$. The derivative $\partial \left(\partial^2 P(\cdot)/\partial A \partial U\right)/\partial U = 0$ implies that the marginal effect of sick absence on penalty if unemployment increases is unaltered if unemployment becomes even higher. Setting $\partial \left(\partial^2 P(\cdot)/\partial A^2\right)/\partial U = 0$ means that the convexity of the penalty function in absence is the same irrespective of unemployment level. All the analytical results can be obtained from the authors upon request.

The reason is that a "low" level of unemployment requires a relatively small reduction in absence to keep penalty constant when unemployment increases. On the other hand, if the level of unemployment is "high", an increase in unemployment "punish" relatively much more so a larger reduction in absence is necessary to keep penalty constant.

By the same line of reasoning, a concave penalty function gives a convex relationship between absence and unemployment. In this setting an increase in unemployment "punish" more if the level of unemployment is "low" as compared to an increase in unemployment when the unemployment level is "high", with corresponding "large" and "small" reductions in absence to keep penalty constant.

In the following we assume that the penalty function is convex in unemployment because workers having high rates of absence are more likely to be dismissed than those with low rates, and because the probability of being dismissed increases with unemployment. In addition, it is much more difficult to get a new job if unemployment is high. Thus it is reasonable that penalty at the margin is higher when the unemployment level is high. The concave relationship between absence and unemployment is illustrated in Figure 1.

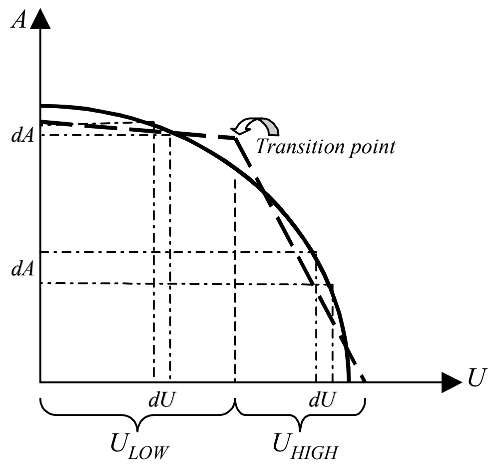


Figure 1: The relationship between absence and unemployment when penalty is convex in unemployment.

The composition hypothesis states that the level of unemployment affects composition of those who have work, and therefore sickness absence: If firms have to increase employment they have to hire workers who have more health problems if unemployment is low.[8] Consequently, the rate of sickness absence is higher when the unemployment level

[8] The underlying assumption here is that firms have full information regarding the distribution of productivity among workers, and that hiring is made according to descending productivity.

is low, and *vice versa*. This implies that an increase in unemployment in a situation with "low" unemployment will reduce sickness absence, whereas an increase in unemployment given a "high" unemployment level will have a much smaller (negligible) effect on absence through composition mechanisms.

On this background we therefore have the following effects, assuming that penalty is convex in unemployment: In a setting of a sufficiently low level of unemployment increased unemployment affects absence through discipline effects (DE in the following) but negligibly through composition effects (CE in the following). A sufficiently high level of unemployment gives the opposite picture, negative DE and negligible CE. In both settings we get a negative total effect. The effects are listed in Table 1.

Estimating unemployment effects which are zero in both settings implies that there are neither composition nor discipline effects. If the estimated unemployment effects are negative in both settings, it follows that the estimate corresponding to "low" unemployment captures CE, whereas the estimate corresponding to "high" unemployment represents DE.

The line of reasoning in Table 1 may also be used to illustrate that a penalty function which is concave in unemployment gives ambiguous predictions regarding CE and DE. In that case DE will be much larger in absolute value when unemployment is "low" as compared to a "high" level of unemployment. But as already explained, the same applies to CE. Hence, it is not possible to separate the two effects because both DE and CE give the strongest effects when unemployment is low.

To illustrate the close connection between our hypotheses and the chosen empirical approach, we have in Figure 1 drawn a linearized version of the concave relationship between absence and unemployment, consisting of two linear parts. The point where these two lines meet gives the transition point which we are going to estimate by applying the Smooth Transition Regression (STR) approach.[9] This approach does not imply that we impose restrictions on the empirical models of a concave relationship between sickness absence and unemployment.

[9] One transition point reflects that we have a logistic STR model in mind, see the Section 4.

Table 1: Identifying discipline and composition effects when the penalty function is convex in unemployment.

Type of effects	Level of unemployment "Low"	"High"
Discipline (DE)	Negligible	Negative
Composition (CE)	Negative	Negligible
Total	Negative	Negative
Estimated effect is zero	No composition	No discipline
Estimated effect is negative	Composition only	Discipline only

3 Data

The analyses are based on aggregate quarterly data collected and published by Confederation of Norwegian Business and Industry (NHO[10]) from the first quarter of 1971 to the second quarter of 1998. NHO is the main organization for employers in manufacturing industries, services and crafts in Norway. Our data refer to blue collar adult workers, separated by gender. The series of absence rates[11] are collected from a constant sample of representative NHO firms. The last included quarter (second quarter of 1998) contained data from firms employing 80 179 workers of which 65 620 were men. Wage data are collected from all member firms of NHO. In second quarter of 1998, wage data were collected from firms with 28 000 female workers and 144 000 male workers.[12]

Duration of absence spells is divided into two categories. The first is short term absence, which refers to sickness spells lasting until 3 days plus absence registered as shirking, hereafter denoted STA. The second category is long term absence, corresponding to sickness spells lasting 4 days or more, hereafter referred to as LTA. This division of short and long term absence spells makes this data set very appropriate for testing our hypotheses as STA has to be absence not covering serious sickness. In relation to

[10] *Næringslivets Hovedorganisasjon.*

[11] Rate of absence= $\frac{\text{registered days absent from work due to sickness}}{\text{number of days actually worked+registered days absent}} * 100.$

[12] See NHO (1971-1998).

possible effects of changes of days without sickness allowances and self-certification in the Norwegian sick pay scheme, this is also a fruitful division.[13] The absence series and the unemployment rate are illustrated in Figures 2-4.

Figure 2 shows that STA is higher among women than men, but the difference becomes smaller throughout the period. In the first eight years of the period, male absence was growing from about 70% to 85% of female absence. From late seventies to mid eighties male absence varied from 95% to about 80% of female absence. From the second quarter of 1986 to 1991 male absence fell more than female absence and in the third quarter of 1990 male absence was about 77% of female absence. The differences were decreasing in the rest of the period, and in 1997 short term absence was approximately at the same level for both genders. However, the two series seem to have similar paths: increases in the seventies, peaks in the early eighties, reductions until third quarter of 1991 and thereafter some increase.

Figure 3 illustrates the paths of LTA for women and men, and the picture is quite different from the STA series.

It is interesting to note that the gender gap does not shrink during the period but on the contrary seems to increase. Male absence was about 75% of female absence in the seventies, but the gap increased constantly throughout the period, and in the end of the period LTA for men was about 60% of female absence. However, the main patterns are quite gender unspecific. LTA fell until 1976-77 and raised to an all time high in 1986/87. Then the absence rates fell again until minima were reached in the mid-nineties. Since then LTA has increased for both women and men.

Figure 4 gives the development in the national unemployment rate. In the seventies the unemployment rate was low and stable, although a weak positive trend is visible. A local peak is reached in 1984, thereafter a rather sharp decrease for a few years only to grow steadily through the rest of the eighties and into the nineties. Then the economy began to grow and the unemployment rate fell into the millennium.

When we compare these series, there are some important features to mention. When unemployment is low but slightly increasing, as in the seventies, LTA is high but decreases slightly, while STA is high and increases. When unemployment increases rapidly, as from 1988, LTA decreases.

[13] For details, see Section 4.2.

Figure 2: Short Term Absence Rates for Women and Men

Quarterly Data and Moving Average

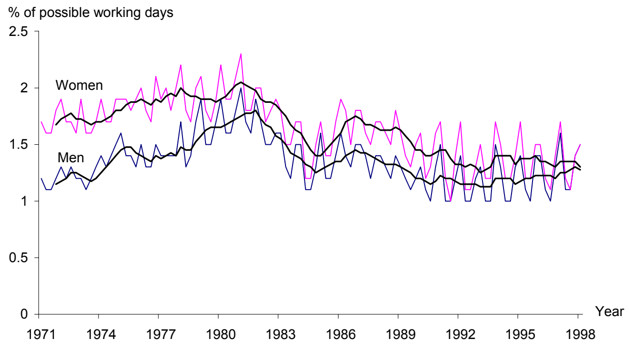


Figure 3: Long Term Absence Rates for Women and Men

Quarterly Data and Moving Average

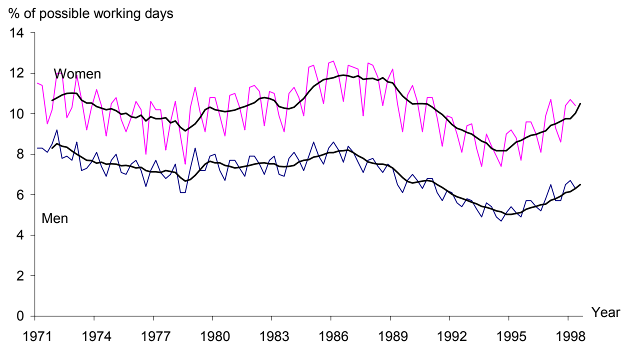
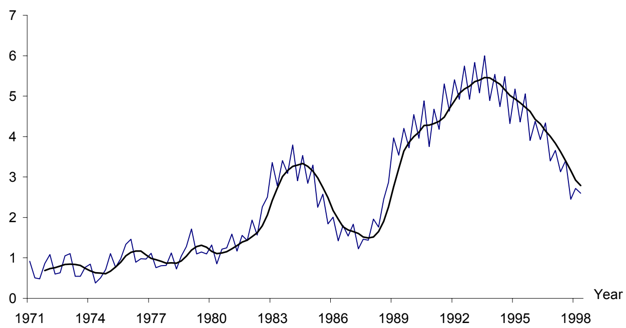


Figure 4: The National Unemployment Rate (per cent)

Quarterly Data and Moving Average



Also STA decreases from 1987. This may indicate different relations between absence and unemployment in periods of high versus low levels of unemployment. However, LTA seems to have a clear inverse relation to the unemployment level while the relation for STA seems to be more mixed. Altogether, this indicates that it is interesting to investigate the relationship between absence and unemployment.

4 Empirical Analysis

We apply the Smooth Transition Regression (STR) model formulated by Teräsvirta (1998) to test the hypotheses outlined in Section 2. This model may in general be written as

$$y_t = x_t'\varphi + (z_t'\theta)G(\gamma, c; s_t) + \varepsilon_t, \tag{3}$$

where x_t is the vector of explanatory variables in the linear part of the model, and z_t is the vector of explanatory variables entering the non-linear part. In our general models all the continuous variables included in the linear part will also be included in the non-linear part (except the dummy variables), so $z_t \in x_t$. The vectors $\varphi = (\varphi_0, \varphi_1, ..., \varphi_p)'$ and $\theta = (\theta_0, \theta_1, ..., \theta_p)'$ contain the parameters of the linear and non-linear parts, respectively, and $\{\varepsilon_t\}$ is a sequence of independent, identically distributed errors. In (3) $G(.)$ is a bounded continuous transition function, and it is customary to bound $G(.)$ between zero and unity. The variable s_t is the transition variable, c is the transition point and γ gives the slope of the transition function. The parameters c and γ will be estimated.

The objective of this paper is to analyse to what extent the unemployment level influences absence, so the rate of unemployment is the obvious choice of transition variable. Furthermore, we have to make a choice of transition function, and one alternative is the logistic STR model (LSTR1) given by

$$G_1(\gamma, c; s_t) = [1 + exp\{-\gamma(s_t - c)\}]^{-1}, \qquad \gamma > 0. \tag{4}$$

This function is monotonically increasing in s_t with $\gamma > 0$ as identifying restriction. The slope parameter γ indicates the speed of transition from zero to unity, and the parameter c locates the transition. If $\gamma \to \infty$ in (4) the STR model becomes a switching regression model with s_t as switching variable. For all four types of sickness absence (cf. Section

3) we start by first estimating rather general STR models and simplify these to well specified parsimonious models through a reduction process. We use the LSTR1 model in this reduction process. However, the validity of using this functional form is checked by testing the parsimonious models against two alternative transition functions.[14]

Referring to the STR model given by equation (3), y_t, and the vectors x_t and z_t are defined as follows:

$$y_t = \Delta a_{ijt} \tag{5}$$

$$\begin{aligned} x_t' &= (1, a_{ijt-1}, w_{it-1}, u_{t-1}, s74_{t-1}, s78_{t-1}, s84_{t-1}, s91_{t-1}, \\ &\quad \Delta a_{ijt-\ell}, \Delta w_{ijt}, \Delta w_{ijt-1}, \Delta u_t, \Delta u_{t-1}, \Delta s74_t, \Delta s78_t, \\ &\quad \Delta s84_t, \Delta s91_t, Q_1, Q_2, Q_3) \end{aligned} \tag{6}$$

$$\begin{aligned} z_t' &= (1, a_{ijt-1}, w_{it-1}, u_{t-1}, \\ &\quad \Delta a_{ijt-1}, \Delta w_{ijt}, \Delta w_{ijt-1}, \Delta u_t, \Delta u_{t-1}) \end{aligned} \tag{7}$$

Respectively, the vectors x_t and z_t correspond to the linear and non-linear parts of our starting general STR models. All but the dummy variables are log transformed, and Δ denotes first difference. The variable a_{ijt} is the rate of sickness absence, where index i refers to gender (w=women, m=men), j refers to duration of absence spells, and t refers to time (quarters), respectively. As explained in Section 3, duration of absence spells is divided into two categories: Sickness spells lasting until 3 days plus absence registered as shirking [a_{w3t} and a_{m3t} in equation (5)], and sickness spells lasting 4 days or more [a_{w4t} and a_{m4t} in equation (5)]. Thus, and consistent with the notation in Section 4, in the following short term absence (STA$_i$) refers to j=3 and long term absence (LTA$_i$) to j=4. The other variables are defined as follows: w_{it} is nominal hourly earnings deflated by the national consumer price index; u_t is the national rate of unemployment; h_t is normal working hours; and Q_r represent seasonal dummies where r refers to corresponding quarter.

The dummy variables sY_t ($Y = 74, 78, 84$ and 91) are included to capture important changes in the sick pay scheme, and take the value 1 from the quarter of change, and

[14] The alternative transition functions are the non-monotonic three-regime STR model and the exponential STR model, respectively $[1 + exp\{-\gamma(s_t - c_1)(s_t - c_2)\}]^{-1}$ (LSTR2) and $[1 + exp\{-\gamma(s_t - c)^2\}]^{-1}$ (ESTR). See Teräsvirta (1998).

zero otherwise. In 1974 the number of days without sickness allowances was reduced from three to one, so that employees receive pay for all days in a sickness spell except the first one. An important reform was implemented 1 July 1978. The reform directed all employers to pay sickness allowances for the first fourteen days of a sickness spell while the rest of the spell should be covered by National Insurance, implying that all groups of employees were given the same formal rights in case of sickness. In addition, the one day without sickness allowances was abolished, and so-called self-certification was introduced, saying that an employee may have until six spells of sickness absence of up to three days duration during the year without being medically certified. In 1984 the total number of self-certifications during the year was reduced from six to four, and the counting of absence days became more restrictive. Several means were introduced in 1991 in order to reduce growth in National Insurance expenditures, of which the most important seem to be to stimulate combinations of pay from work and sickness allowances, and more effective control of doctors' practice of issuing medical certificates. However, the 1991 dummy is also included to capture possible effects of the so-called LO-NHO project to reduce sickness absence.[15]

The parameters in the STR model are estimated by applying a variable-metric approach algorithm for non-linear optimization according to the BFGS update (see Hendry, 1995). Initial parameter values may be critical for reaching global maxima, and we have used estimates obtained from OLS regressions for different sets of γ and c as such values. Combinations of estimates giving the smallest sum of squared residuals are used as initial values.[16]

Parameter estimates of the parsimonious STR models are given in Appendix, and these models have been exposed to several mis-specification tests.[17] None of the null hypotheses of no autocorrelation in the error terms are rejected at a 5% level of significance (LM tests). We also conclude that the models are well specified with respect to functional form of the transition function, and that unemployment as transition variable models non-linearity sufficiently in all cases.[18] Following the test suggested by Eitrheim and

[15] This project started in 1991, and is a co-operation between Norwegian Confederation of Trade Unions (*LO*) and Confederation of Norwegian Business and Industry (*NHO*).

[16] Usually 100 $\gamma's$ and 50 c's are tried, *i.e.*, 5,000 different OLS regressions are compared.

[17] The detailed test results can be obtained from the authors upon request.

[18] The STA_M model is not tested because all the variables (with the exception of dummies) entering the linear part of the model also are included in the non-linear part. Thus there are no variables left to

Teräsvirta (1996), we find no indications of remaining non-linearity in the models for STA and LTA_M. However, there are indications of the opposite conclusion regarding LTA_W. Potential remaining non-linearity in LTA_W has been investigated by including alternative transition variables in the LTA_W model. Due to increased complexity both regarding interpretation and testing these specifications have been discarded. We find no statistically significant indications of non-constant parameters in the models for STA. Considering the non-constancy tests of LTA_W, one or more of the non-dummy parameters in the linear part of the model seems to be linearly trended.[19] From the tests of LTA_M, we find indications of a non-constant intercept in the non-linear part of the model.

The STR methodology requires stationary variables, and this also applies to the transition variable. Augmented Dickey Fuller tests indicate that all the level variables included in our models have unit roots, also the rate of unemployment. This means that we by assumption consider unemployment to be a stationary variable, which is a reasonable assumption on theoretical grounds. To test whether the level variables co-integrate, we have applied an augmented Dickey Fuller test on the residuals obtained from the parsimonious STR models. Based on this test we conclude that the level variables co-integrate.[20]

The above referred mis-specification tests tell us that the models are fairly well specified. However, we find indications of parameter non-constancy in the LTA relations, but the problem is first and foremost related to LTA_W. For LTA_W we also find indications of remaining non-linearity. On the other hand, the STA models pass all tests.

5 Results

In Tables 2 and 3 the long and short run parameter estimates of the parsimonious STR models are given for unemployment levels sufficiently below and above the points of transition, so that the values of the transition functions equal zero and one, respectively.[21] When commenting the results we refer to G=0 as "low unemployment", which means that unemployment is sufficiently below the estimated transition point to give this value of the estimated transition. Consequently, G=1 refers to "high unemployment", which implies

be included in this second transition function.

[19] After a more thorough testing sequence we find indications that non-constancy may be related to the parameters of a_{t-1} and Δa_{t-1}.

[20] This conclusion is based on critical values for linear models.

[21] The parameter estimates and corresponding t-values of these four models are given in Appendix.

that unemployment is sufficiently higher than the estimated point of transition.[22]

From Table 2 it is seen that for short term absence among women, STA_W, the estimated point of transition is 2.6%. Unemployment was below this level until 1983. After a short period of relatively high unemployment, the rate fell below 2.6% in the beginning of 1985, but from 1989 above 2.6%. (See Figure 3) We also see that the transitions occurs relatively fast since γ is estimated to 11.5. For STA men, STA_M, the transition point is estimated to 1.7%. The rate of unemployment was below this level until 1982 and the two years 1986 and 1987, and above during the years 1982-85. From 1989 and throughout the estimation period it was above. As γ is estimated to 484, the STR model for STA_M clearly is a switching regression model. The difference between the transition points for STA_W and STA_M seems significant. However, the periods of unemployment above and below these transition points are not very different because unemployment rises and falls sharply in the actual intervals of transition, cf. Figure 3.

The points of transition in long term absence also differ for women and men, as they are estimated to 1.6% for LTA_W and 1.9% for LTA_M. The difference is not as large as for STA, so the periods of unemployment above and below the points of transition are very much the same. A more important difference concerns the estimate of γ, which is much lower for LTA_W, giving a much smoother transition than for men.

In Table 2 and 3 we have also reported results from linear models, denoted LIN. These models are very similar to the models estimated by Dyrstad and Lysø (1998). The models in *ibid.* are estimated on data from the period 1971-92, while the LIN models are estimated on the same sample as is used for the STR models [1971(3)-1998(2)]. The estimates are very similar with a few exceptions, of which one is particularly important in relation to the topic of this paper: The estimated unemployment elasticities for LTA are much larger in absolute value in the LIN models as compared to those estimated on the 1971-92 sample. Furthermore, the wage rate and the 1991 dummy in the LTA_M relation have statistically insignificant effects in the LIN model. Mis-specification tests indicate that the models in *ibid.* and the LIN models are well specified. Based on estimated standard errors the STR models improve fit by approximately 15% for STA and 10% for LTA relative to the LIN models.

[22] As explained in a preceding section, the transition from G=0 to G=1 also depends on the size of γ, so when we use the terms "sufficiently above" or "below", we mean that unemployment is so high that we are in one of the corresponding states.

Table 2: Parameter Estimates from STR and Linear Models (LIN)

Short Term Absence (STA)

	Women			*Men*		
	G=0	*G=1*	*LIN*	*G=0*	*G=1*	*LIN*
Long Run Estimates						
w	0.34	0.23	0.38	-	-	-
u	-	-	-0.13	-0.13	-0.19	-0.17
s74	-	-	-	0.17	0.08	0.19
s78	-	-	0.06	0.24	0.12	0.23
s84	-0.17	-0.12	-0.15	-0.24	-0.12	-0.15
s91	-0.17	-0.12	-0.15	-	-	-
Short Run Estimates						
Δw_t	-	-	-	-	-	-
Δu_t	-0.10	-0.10	-0.18	-	-0.15	-0.19
Δa_{it-1}	-0.23	0.21	-	-0.25	0.36	-
Δ *s84*	-0.20	-0.20	-0.24	-0.17	-0.17	-0.17
Intercept	-1.50	-1.50	-1.67	0.15	0.59	0.18
Q_1	0.07	0.07	0.05	0.09	0.09	0.08
Q_2	-0.14	-0.14	-0.21	-0.06	-0.06	-0.19
Q_3	-0.13	-0.13	-0.14	-0.12	-0.12	-0.14
γ		11.53	-		484	-
c		2.57	-		1.65	-
R^2		0.89	0.84		0.90	0.86
*100*SSE*		5.65	6.62		5.46	6.25

*NOTES: R^2=coefficient of determination, 100*SSE=estimated standard error (in per cent).*

There are important differences to be noted when we compare the STR results and the LIN results, and when we compare results from periods of high and low unemployment obtained from the STR models. In the following we concentrate on discussing unemployment effects, but also results concerning wage changes and changes in the sick pay scheme will be discussed. We start by first looking at the results for STA.

Based on the STR model unemployment does not affect STA_W in the long run, which is contrary to the estimate from the LIN model. On the other hand, for STA_M both the STR model and the LIN model indicate that unemployment is important. But the differences between high and low unemployment, and between the STR model and the LIN model are small. These results indicate that in the long run neither discipline (DE) nor composition (CE) are important for STA_W, while both DE and CE may be important for STA_M. The short run unemployment elasticities for STA_W are significant but identical in high and low unemployment, and smaller in absolute value than the estimate obtained from the LIN model. For STA_M we estimate only short run effects when unemployment is high, and this estimate has approximately the same size as in the LIN model. According to our hypotheses in Section 2 these results could be interpreted as short run DE for STA_M, and a mixture of short run DE and CE for STA_W.

STA_M is not affected of wage changes in the long run, and there are no short run wage effects present in short term absence, irrespective of model. But there is a difference regarding women: The long run wage elasticity is higher when unemployment is low, whereas the corresponding estimate in LIN is marginally higher.

Turning to effects of changes in the sick pay scheme, we estimate both short and long run effects of the 1984 reform. The short run estimates are identical in high and low unemployment, and rather similar both in the STR and LIN models. The more generous changes in 1974 and 1978 increase STA_M to a much larger extent in periods of low unemployment. The long run elasticities are respectively .17 and .24 in low unemployment, but only the half when unemployment is high. Sick pay scheme changes could be interpreted as discipline mechanisms because they change employees' budget restrictions, so these differences are reasonable. With estimates of -.24 in low and -.12 in high unemployment the same applies to the restrictions implemented in 1984: In periods of low (high) unemployment the discipline effects are weaker (stronger), so there is more (less) "room" for absence reductions due to for instance restrictions in the sick pay scheme. Regarding the 1974 and 1978 reforms the results from the LIN model are very similar to the STR results when unemployment is low, but similar to the results corresponding to periods of high unemployment regarding the 1984 reform.

The reforms in 1974 and 1978 do not affect STA_W, whereas the restrictions implemented in 1984 and 1991 affect this type of absence slightly more in periods of low unemployment. It follows from Table 2 that these results are very similar to those obtained from the LIN model.

In the sense that the unemployment estimates indicate stronger discipline effects among men than among women, the above interpretations of the reform effects are consistent with the interpretations of the unemployment effects.

Turning to long term absence in Table 3, there are also some important differences both between the STR models and the LIN models, and between high versus low unemployment within the STR models.

The long run effects of unemployment changes from the STR model are larger in absolute value than the LIN results, and we also see that LTA is only affected during periods of high unemployment. It is also interesting to note that the long run elasticities are rather similar in size, -.41 and -.48 respectively for women and men. The short run elasticities are much smaller, and similar in periods of high and low unemployment for STA_W. A short run effect is obtained only in periods of high unemployment for LTA_M.[23] Within the framework of our hypotheses in Section 2, these results have to be interpreted as strong discipline effects.

Wage changes also affect long term absence differently in the STR and LIN models, and there are large differences between low and high unemployment. In absolute value the long run wage elasticities are half the size in periods of low unemployment as compared to periods of high unemployment. For LTA_W the estimate from LIN is approximately equal to the STR estimate corresponding to low unemployment, whereas wage changes do not affect LTA_M in the LIN model. Short run wage effects are only obtained from the STR model for LTA_M with no difference between low and high unemployment.

We estimate negative wage elasticities, which within the labour supply model in Section 2 should be interpreted as dominating substitution effects. The positive estimates of the wage elasticities for STA_W indicate dominating income effects for this type of model.

We also obtain important results regarding changes in sick pay scheme. The long run effects of the 1978 change are much smaller in the STR model than in the LIN model, and there are differences between low and high unemployment. For both women and men the estimates in high unemployment are not far from twice the size of the estimates in low unemployment. We expect these estimates to be positive, so it is surprising that the short run impact effect in the STR model for LTA_W is negative. This could be explained as adjustment effects in wake of the long run increase in absence.

The change implemented in 1984 affects only LTA_W, and the long run estimates are smaller when unemployment is low than high, whereas the estimate from LIN is the average of these two estimates. Short run effects of the 1984 change are also estimated for LTA_W only, and there is no difference between low and high unemployment. As the 1984 reform implied a more restrictive sick pay scheme directed towards STA, we would not expect this reform to affect LTA.

[23] The short run unemployment dynamics in the LIN model includes $\Delta^2 u_t$, saying that the speed of change in unemployment is important. This variable was included in Dyrstad and Lysø but in the STR model we assume the non-linear components to be sufficiently modelled by the transition function.

Table 3: Parameter Estimates from STR and Linear Models (LIN)

Long Term Absence (LTA)

	Women			Men		
	G=0	G=1	LIN	G=0	G=1	LIN
Long Run Estimates						
w	-0.35	-0.70	-0.31	-0.51	-0.95	-
u	-	-0.41	-0.12	-	-0.48	-0.33
s78	0.07	0.11	0.21	0.08	0.14	0.38
s84	0.11	0.18	0.15	-	-	-
s91	-	-	-0.09	-	-	-
Short Run Estimates						
Δw_t	-	-	-	-0.38	-0.38	-
$\Delta^2 u_t$	-	-	-0.10	-	-	-0.04
Δu_t	-0.07	-0.07	-	-	-0.13	-
Δa_{it-1}	-	-0.28	-0.16	-0.18	-0.18	-0.29
Δ *s78*	-0.06	-0.06				
Δ *s84*	0.07	0.07	-	-	-	-
Intercept	3.77	3.77	1.68	2.5	2.25	0.26
Q_1	-	-	-	-	-	-
Q_2	-0.14	-0.14	-0.18	-0.14	-0.14	-0.15
Q_3	-0.22	-0.22	-0.19	-0.12	-0.12	-0.13
γ		2.47	-		20.5	-
c		1.64	-		1.86	-
R^2		0.93	0.92		0.87	0.82
*100*SSE*		3.56	3.93		3.19	3.60

NOTES: See Table 2.

Dyrstad and Lysø (1998) get a similar contradiction, and they point at three possible explanations. First, the restrictions implemented in 1984 are first and foremost expected to reduce STA because the number of self-certifications of up to three days duration is reduced, which directly affects STA. The results in Table 2 confirm this expectation. One implication of this could be that workers instead of staying home to get well, they go to work so their health conditions worsen, ending up in LTA spells. The result is higher LTA. Secondly, a reduced number of self-certifications means that it becomes more difficult to be absent from work, so some people may simply choose longer absence spells if they have to be absent. Third, the more restrictive way of counting absence days, which also was a part of the 1984 reform, may give a direct increase in LTA: If a person is absent from work on a Friday and the next Monday, the number of registered absence days before the reform is two days and after four days. In this example the immediate effect is that STA is reduced by two days whereas LTA is increased by four days. According to our results all three explanations are possible. Even if the reductions in STA_W - which are approximately 17% and 12% in low and high unemployment - are directly transferred into higher LTA_W because of change in counting of absence days, these reductions are simply to small to match the 11% and 18% increase in LTA_W.[24]

On the data set used in this paper only the LIN model for LTA_W gives a statistically significant estimate of the 1991 dummy. This is also an important result because it does not support previous conclusions that the LO-NHO project reduced absence by approximately 20%.[25] Thus the results from the linear model estimated on this longer sample period, together with the STR results, indicate that the effects of this project is over-estimated. As unemployment increased to much higher levels in the 1990s than in the 1970s and 1980s, it seems that the results previously explained by the 1991 changes are due to high unemployment.

6 Conclusions

In this paper we have applied the Smooth Transition Regression (STR) approach to analyse non-linear unemployment effects in sickness absence among blue collar workers in Norway in the time period 1971-1998. The estimated STR models seem well specified with the possible exception of long term absence among women. Compared to a well specified linear model, these models also explain more of the variation in absence.

The most important aim of this paper has been to try to identify whether the negative relationship between sickness absence and unemployment is due to discipline or composition effects, or both. We think that applications of STR models have been very fruitful in this respect as a number of new results have appeared.

The estimation results point to the presence of convex penalty functions. For long term absence, the estimated long run unemployment elasticities are -.4 for women and -.5 for men if the rate of unemployment is higher than 1.6-1.9%. These estimates are much larger in absolute value than those reported in previous studies on Norwegian data. However, there are no effects of changes in unemployment on long term absence if the unemployment level is below the transition values. According to our identification hypotheses, these results must be interpreted as discipline effects and not composition effects. On the

[24] The ratio LTA_W/STA_W was in 1984 approximately 7.5, so STA_W has to be reduced by at least 66% and 123% in order to match the increases in LTA_W.

[25] Bru *et al.* (1994) is one evaluation report concluding that the LO-NHO project reduced absence by 20%, and Dyrstad and Lysø (1998) get long run effects supporting this conclusion, at least for LTA_W.

other hand, neither discipline nor composition effects seem to be present for short term absence among women. Short term absence among men is affected by changes in unemployment but the unemployment effects are much smaller as compared to those for long term absence, and it is not possible to draw conclusions regarding the relative importance of discipline and composition effects for this type of absence.

These differences between short and long term absence are opposite to those reported in Dyrstad and Lysø (1998), and are partly explained by the application of non-linear models, and partly by a longer estimation period containing both higher levels of unemployment and larger unemployment changes. Under the reasonable assumption that it is more serious sickness among those on long term absence spells than among those on short term spells, one may argue as in *ibid.*, that only those who are not seriously sick can be disciplined. Hence, one expect discipline effects first and foremost to appear in short term absence. The long term absence rates which we have used in this analysis refer to absence spells lasting from four days to one year. This means that rather short absence spells - but of more than three days duration - are classified as long term absence, and that the above argument may still be valid. Thus both the longer estimation period and the STR approach have revealed other relationships between absence and unemployment, and other interpretations.

One may also put the argument above the other way around: Those with bad health and corresponding long term absence spells are more afraid of loosing their jobs than those on short term absence spells because they see themselves as marginal workers. A consequence of this is that it is not surprising to find the strongest discipline effects in long term absence.

The estimated wage effects support this conclusion. The wage elasticities for long term absence are negative and much larger in absolute value when unemployment is high, which is most reasonably interpreted as due to discipline mechanisms: If unemployment is high, and the wage level increases, it also becomes more expensive to loose a job. In accordance with this, we estimate positive elasticities for short term absence among women with a much smaller difference between high and low unemployment, and find no significant wage effects for short term absence among men.

As mentioned in the introduction, composition and discipline effects have different policy implications: Composition effects point in direction of micro measures whereas discipline effects point at macro measures. Finding that discipline effects are more important than composition effects thus points at restrictions in the sick pay scheme if the government wants to reduce sickness absence. This conclusion is largely supported by the estimated effects of actual changes in the sick pay scheme during the estimation period.

The sick pay scheme changes in 1974, 1978 and 1984 affect first and foremost short term absence with estimates according to our expectations, pointing at discipline effects also for this type of absence. Looking at long term absence, however, the picture is a bit mixed. As the 1974 reform was directed towards short term absence, it is not surprising that it gives no effects on long term absence. The *generally* more generous sick pay scheme implemented in 1978 increases long term absence for both women and men. One may question whether the result that these effects are larger when unemployment is high is consistent with the above interpretations of discipline effects. Our answer is yes, because discipline effects are stronger when unemployment is high, so that the discipline pressure may be reduced due to this general improvement in the sick pay scheme.

Also the restrictions implemented in 1984 were directed towards short term absence, so it is a bit surprising that they increase long term absence among women, and that the increases are stronger if unemployment is high. The 1984 reform reduce short term absence among women more strongly if unemployment is high, so it is possible that this pops up in a stronger increase in long term absence if unemployment is high. Contrary to this is the fact that we do not see this relationship for men, as the 1984 change does not affect long term absence for this group.

With the exception of the 1978 reform, these reforms concern short term absence most directly, whereas the estimated discipline effects are found to be strongest for long term absence. Consequently, and in accordance with our results, policy measures should be directed towards the particular absence spells which one want to affect, and at the same time take possible spill-over effects on the length of absence spells into account.

In general our results point at the importance of keeping an eye on the labour market because both wage and unemployment changes, and particularly the *level* of unemployment, seem to be very important. If unemployment is low, and seems to be staying low, policy measures seem to be less effective as compared to a situation of high unemployment.

References

[1] Allen, S G (1981): “An Empirical Model of Work Attendance”, *Review of Economics and Statistics*, 63, 77-87.

[2] Barmby, T, J Sessions and J Treble (1994): ”Absenteeism, Efficiency Wages and Shirking”, *Scandinavian Journal of Economics*, 96(4), 561-566.

[3] Brown, S and J Sessions (1996): ”The economics of absence: Theory and Evidence”, *Journal of Economic Surveys,* Vol. 10, No 1.

[4] Bru, E, T Lie and H S Larsen (1994): ”Sammen om sykefravær” (”Together on sickness absence”), Report RF 120/94, Rogalandsforskning, Stavanger.

[5] Chaudhury, M and I Ng (1992): “Absenteeism Predictors: Least Squares, Rank Regression, and Model Selection Results”, *Canadian Journal of Economics*, 3, 615-35.

[6] Dahl, S Å (1990): ”Sysselsetting, sykefravær og sykelighet 1980-1995. En makrososiologisk analyse” (”Employment, sickness absence and sickliness 1980-1995. A macro-sociological analysis”), Report, FAFO, Oslo.

[7] Doherty, N A (1979): “National Insurance and Absence from Work”, *The Economic Journal*, 89, 50-65.

[8] Drago, R and M Wooden (1992): “The Determinants of Labour Absence: Economic Factors and Workgroup Norms Across Countries”, *Industrial and Labour Relations Review*, 45, 764-78.

[9] Dyrstad, J M and N Lysø (1998): “Økonomiske faktorer bak sykefraværet” (”Economic factors behind sick absence”), *Norsk Økonomisk Tidsskrift* (Norwegian Journal of Economics), 112, 155-184.

[10] Eitrheim, Ø and T Teräsvirta (1996): “Testing the Adequacy of Smooth Transition Autoregressive Models”, *Journal of Econometrics,* 74, 59-75.

[11] Halvorsen K and A Johannesen (1991): ”Når bedriften ikke har bruk for deg lenger” (”When the firm does not need you anymore”), NKSH, Oslo.

[12] Hem, K-G (2000): "Økonomiske konsekvenser av sykefravær" ("Economic consequences of sickness absence"), Report STF78 A00508, SINTEF Unimed, Oslo.

[13] Hendry, D F (1995): *Dynamic Econometrics*, Oxford University Press, Oxford.

[14] Hippe, J and A W Pedersen (1991): "Førtidspensjonering i private bedrifter" ("Early retirement in private firms"), Report 813, Fafo, Oslo.

[15] Johansen, P and M Palme (1996): "Do Economic Incentives Affect Work Absence? Empirical Evidence Using Swedish Micro Data", *Journal of Public Economics*, 59, 195-218.

[16] Kenyon, P and P Dawkins (1989): "A Time Series Analysis of Labour Absence in Australia", *Review of Economics and Statistics*, 71, 232-39.

[17] Lantto, K and E Lindblom (1987): "Är arbetslösheten hälsosam?" ("Is unemployment healt-bringing?"), *Ekonomisk Debatt*, 333-36.

[18] Næringslivets Hovedorganisasjon, NHO (1971-1998): "Lønns og fraværsstatistikk" ("Wage and absence statistics"), Oslo.

[19] Teräsvirta, T (1998): "Modeling Economic Relationships with Smooth Transition Regressions", in A Ullah and D E A Giles (eds): *Handbook of Applied Economic Statistics*, 507-552, Marcel Dekker Inc, New York.

APPENDIX

Parameter Estimates and t-values of Parsimonious Models

	Short term absence, women	*Short term absence, men*	*Long term absence, women*	*Long term absence, men*
γ	11.53 *1.88*	484.40 *73.73*	2.47 *1.48*	20.52 *0.69*
C	2.57 *22.44*	1.65 *11.05*	1.64 *4.20*	1.86 *26.65*
Constant	-1.50 *-2.71*	0.15 *3.74*	3.77 *6.73*	2.50 *4.56*
a_{t-1}	-0.77 *-7.85*	-0.76 *-4.86*	-0.74 *-8.58*	-0.39 *-5.02*
Δa_{t-1}	-0.23 *-2.13*	-0.25 *-2.07*		-0.18 *-2.95*
w_{t-1}	0.26 *3.55*		-0.26 *-4.38*	-0.20 *-3.69*
Δw_{t-1}				-0.38 *-1.96*
u_{t-1}		-0.10 *-2.47*		
Δu_t	-0.10 *-2.77*		-0.07 *-3.13*	
$s74_{t-1}$		0.13 *4.05*		
$s78_{t-1}$		0.18 *5.63*	0.05 *2.94*	0.03 *2.70*
$s84_{t-1}$	-0.13 *-4.67*	-0.18 *-6.85*	0.08 *4.80*	
$s91_{t-1}$	-0.13 *-5.34*			
$\Delta s78_{t-1}$			-0.06 *-1.58*	
$\Delta s84_t$	-0.20 *-3.19*	-0.17 *-2.85*	0.07 *1.91*	
Q1	0.07 *3.23*	0.09 *4.00*		
Q2	-0.14 *-4.39*	-0.06 *-2.02*	-0.14 *-9.36*	-0.14 *-12.77*
Q3	-0.13 *-6.34*	-0.12 *-5.09*	-0.22 *-18.32*	-0.12 *-10.23*

Non-linear parameters:				
Constant		0.44 *5.20*		-0.25 *-1.54*
a_{t-1}	-0.34 *-6.94*	-0.79 *-5.18*	0.30 *3.01*	0.18 *2.37*
Δa_{t-1}	0.44 *4.70*	0.61 *5.25*	-0.28 *-4.31*	
w_{t-1}			-0.05 *-1.42*	
u_{t-1}		-0.19 *-4.05*	-0.18 *-3.34*	-0.10 *-3.37*
ΔU_t		-0.15 *-2.74*		-0.13 *-4.17*