WORKING PAPER No. 23
February 1993

FORECASTING CAR EXPENDITURES USING HOUSEHOLD SURVEY DATA - A COMPARISON OF DIFFERENT PREDICTORS

Bo Jonsson

Svensk resumé på sid 25.

Uppsala University
Department of Statistics
P.O. Box 513
S-751 20 Uppsala
Sweden

January 1993

FORECASTING CAR EXPENDITURES USING HOUSEHOLD SURVEY DATA

- A COMPARISON OF DIFFERENT PREDICTORS

by

Bo Jonsson

This research reported has been supported by the Swedish Council for Research in the Humanities and Social Sciences (HSFR) and by the National Institute of Econimic Research (KI).
Helpful comments from Anders Ågren, Anders Christoffersson, Lars-Erik Öller and Wladislaw Milo are gratefully acknowledged.

ABSTRACT

Time series data based on surveys contains sampling errors. In this paper the predictive ability of household survey data (concerning attitudes about the economic development and car buying intentions) on consumer expenditure on automobiles in Sweden is studied. This is done for some proportion variables on different levels based on the survey data and thus these regressors contain errors. The paper focuses on a comparison of the OLS predictor, a predictor based on consistent estimation and one based on a modified OLS estimator. It turns out that the OLS predictor is often the best overall predictor, while other predictors sometimes should be used depending on the relation between the error variance of the estimation and prediction periods and the cyclical position of the prediction value of the survey based regressor.

Keywords: Prediction; Errors in variables; Household survey data, Car expenditures.

1. INTRODUCTION

The predictive ability of household survey data, collected over time, on total consumption and on consumer expenditures on durables has been studied in several countries (for surveys, see Pickering 1984, Kamakura and Gessner 1986, and Ågren 1989). The survey data used concerns attitudes towards the general and private economic conditions and buying plans. The emphasis of the studies is on the predictive ability of models where ordinary macro economic variables and attitudes/plans are taken together and models where only attitude/plan variables are included. In a recent paper Jonsson and Ågren (1992) examine the usefulness of such survey data as predictors of the development of household expenditures on automobiles in Sweden. They concentrate on the latter type of models since their interest focuses on quick indicators that require data that are easier to produce than national accounts' data and hence are available without long delays. In their investigation Jonsson and Ågren considered the explanatory power as well as the prediction accuracy of such indicators. The results showed that the indices mainly contained information about the quarter in which they were collected and that the best single indicator is to be found among the plan indices. However, an indicator based on car registration statistics turned out to be at least as good. By combining plan/attitude indices with car registrations the study indicated that substantial improvements can be obtained as regards both explanatory power and prediction accuracy.

It was also found that the choice of the best performing index was dependent on the sample sizes of the surveys. This is because the indices are sample estimates of population quantities and thus contain sampling errors. If observations corresponding to quarters with small surveys were excluded, the explanatory power increased for some variables and remained relatively unchanged for others. Higher coefficients of determination were obtained especially for proportion variables on low levels as, for instance, the percentage of households that are 100% certain of buying a new car within six months.

The explanatory power of a single regression model is under usual assumptions of independence proportional to the so called reliability ratio

$$K=\sigma^2_x/(\sigma^2_x+\sigma^2_u), \qquad (1)$$

where σ^2_x is the variance of the true regressor (x) and σ^2_u is the error variance in the observed regressor (X) (see e.g. Fuller, 1987 p.4). It is also well known that the probability limit of the OLS estimator of the slope parameter β is $K\beta$. These results extend to the case of unequal error variances, where σ^2_u is the mean error variance. The effect of sampling errors on the estimation of β in the situation of a regressor consisting of population proportions estimated through surveys based on simple random sampling was studied in Jonsson (1992a). He showed that the effect depends primarily on three factors, the size of the survey, the coefficient of variation of the true regressor and its level. Given the first two factors it was found that the inconsistency of OLS was very sensitive to the level and could be very serious for variables close to zero. Jonsson (1992a) also studies the performance of some consistent moment estimators. When there is considerable OLS bias those estimators outperform the OLS estimator in terms of mean square error (MSE) .

Yet, when it comes to prediction, OLS should be used in the case of stochastic x if the distribution of x and the error generating process are both the same for the estimation and the prediction period (see e.g. Fuller 1987, pp 75). Jonsson (1992b) found this often to be the case, even in the non-stochastic case. He studied the MSE of different predictors based on observed X for different values of the true x and found that the true value of the regressor for which prediction is to be made has to lie rather far away from the mean of the regressor in the estimation period for a predictor based on consistent estimation to be better than the OLS based predictor. If the error variance of the regressor is larger in the prediction period than its mean error variance in the estimation period this interval becomes wider. A modified OLS predictor, taking into account differences in error variances, was sometimes found to yield better predictions than the OLS predictor.

Looking at the historical outcomes of an index it may be unreasonable to assume that the next observation is a random selection from the population that generated the estimation period sample. We often know from the most recent observations where in the business cycle we are. Furthermore the error variances cannot be assumed constant because of unequal sample sizes of the surveys and varying levels of the series. This may motivate the use of other estimators than OLS. The purpose of this study is to extend the study of Jonsson and Ågren (1992) with a predictor based on consistent estimation of the parameters and a modified OLS predictor that takes into account differences in error variances between the prediction and the estimation periods. We then compare the forecast performance of these predictors and the OLS predictor. This will be done for some of the indices used by Jonsson and Ågren.

2. MODELS AND METHODS

2.1 Model. Jonsson and Ågren (1992) used a model where the variable logged car expenditures (*CA*) was linearly related to a single index and seasonal dummies. A preliminary analysis and an underlying assumption of a multiplicative seasonal factor supported the choice of using logged car expenditures. They also tried models with several indices. Among those was a model including as a regressor the number of registrations of new cars (*CREG*) in the first month of each quarter. Data on this variable are available at about the same time as the result from the household survey carried out by Statistics Sweden. Studies were also performed using lagged indicators but since the indices were found to primarily give information about the current quarter we will not consider lagged indicators. Furthermore we will restrict the models to include only one attitude/plan variable. Hence the models to be used here are:

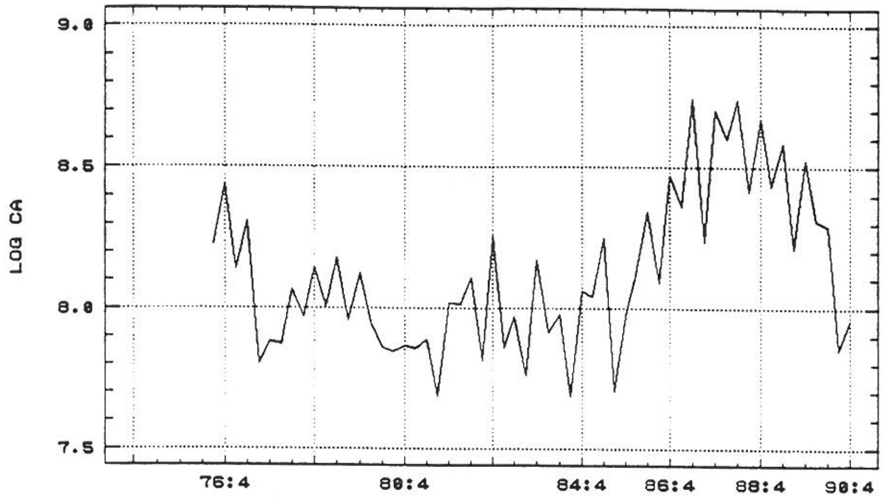
Model A: $$ln\ CA_t = \alpha_0 + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \beta_1 x_t + \varepsilon_t \qquad (2A)$$

Model B: $$ln\ CA_t = \alpha_0 + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \beta_1 x_t + \beta_2 CREG_t + \varepsilon_t\, , \qquad (2B)$$

where $X_t = x_t + u_t$ is observed instead of x_t with error variance σ^2_{ut}. D_{it} denotes the seasonal dummy variable for the $i^{th.}$ quarter, x_t the population value of an index and X_t the survey value of the index at time point t. The model specification implies that only the attitude and plan variables are assumed to be measured with errors.

2.2 Method. We will carry out the forecast evaluation in the same way as in Jonsson and Ågren (1992). There the forecast evaluation was based on the period 1985:1-1990:4. In Figure 1 we have illustrated the development of consumer expenditures on cars from 1976:3 - 1990:4. As can be noted, the forecast period includes a boom in car expenditures during the late eighties with a period of increase (1985-86), high-level expenditures (1987-88) and decline (1989-90). The models will be reestimated successively starting with the period 1976:3 to 1984:2 making a forecast for 1985:1 and then adding observations one by one. The reason why the estimation period ends three quarters earlier is the delay in the production and dissemination of car expenditure data.

FIGURE 1. Logged car expenditures during 1976:3-1990:4



2.3 Large sample bias of OLS when only one regressor is measured with error. Although we primarily are interested in prediction it can be worth discussing the effects of the error in the index variable on the parameter estimates. The effect on the slope parameter in a simple regression model is related to the reliability ratio according to (1). Under general assumptions for a multiple linear regression model where all true variables are measured as deviations from means, the probability limit of the OLS estimator can be shown to be (see e.g. Levi 1973):

$$\text{plim}\,\hat{\boldsymbol{\beta}}_{OLS} = (\boldsymbol{\Sigma}+\boldsymbol{\Omega})^{-1}\boldsymbol{\Sigma}\boldsymbol{\beta} \quad (3)$$

if the true regressors and their errors have finite limiting variance-covariance matrices **Σ and Ω**, respectively. Models A and B have several regressors of which only one is measured with error. By applying a matrix identity[1] due to Bartlett (1951) on (3) it follows that the probability limits of the OLS estimator are

$$\beta_k^* = \frac{\sigma_k^{2*}}{\sigma_k^{2*} + \sigma_u^2}\beta_k, \text{ and} \tag{4A}$$

$$\beta_i^* = \beta_i + \frac{\sigma_u^{2*}}{\sigma_k^{2*} + \sigma_u^2}\gamma_i\beta_k \qquad \forall\, i \neq k. \tag{4B}$$

In (4) the subscript k is used to denote the regressor measured with error. $\sigma^{2*}{}_k$ is the error in equation variance when taking the regression of true x_k on the remaining regressors (x_i, $i \neq k$). $\beta^*{}_i$ is the probability limit of the OLS estimator of the parameters of the regressor x_i and γ_i is the parameter of x_i in the mentioned regression for x_k. If no variation in x_k can be explained by the other regressors, then (4A) yields the same result as that obtained for a simple linear regression model (see Section 1) and the parameters of the other variables are consistently estimated. If the regressors measured without error can explain some variation in x_k, the bias in β_k is increased. However, the other parameters are overestimated (if γ_i and β_k are positive), which may reduce the effect of errors in variables when making predictions. If a large part of the variation in x_k is explained by the other variables, the latter take over almost completely and the probability limit of the OLS estimator of β_k is close to zero. The results in (4) can also be shown to cover the case of non-stochastic regressors and unequal error variances between the fixed values of x_k. $\sigma^2{}_u$ is then the mean error variance in X_k. The equations in (4) were first demonstrated in Chow (1957, pp.94-

[1] $(A + cbb')^{-1} = A^{-1} - \frac{c}{1 + cb'A^{-1}b}A^{-1}bb'A^{-1}$, where A is a symmetric nonsingular matrix, b is a vector and c is a scalar. In our case let A=Σ, c=$\sigma^2{}_u$ and b a vector of zeros except for element k which is set to one.

98). He assumes a classical normal errors-in-variables model, and hence (3) and (4) concern expectations.

3. DESCRIPTION OF THE INDICATORS

3.1 The Household Survey. In October 1973 Statistics Sweden started a regular quarterly survey on consumer attitudes and buying intentions. Since the start, the sample size and sample design have undergone several changes . The sample size is presented in Table 1 and as can be noted it has been reduced over time. In the beginning the number of households investigated was about 10000. From July 1985 on, the January and July surveys cover 1500 households and the April and October surveys 4200 households. The two panels are independent and one third is replaced at every occasion. During the period October 1973 to July 1984 one panel was used and one fifth was replaced at each occasion. One consequence of the panel design is that the errors of an index may be correlated. The sample scheme is stratified sampling with larger inclusion probabilities for high than for low income households. The proportion of nonresponse varies from 11% to 23%. The highest values occur at the July surveys.

Table 1. Sample sizes used in the household survey.

Period	January	April	July	October
1973	-	-	-	10 000
1974-1978	10 000	10 000	10 000	10 000
1979	10 000	10 000	6 600	6 600
1980-1983	6 600	6 600	6 600	6 600
1984	6 600	6 600	6 600	1 500
1985	6 600	1 500	1 500	4 200
1986-	1 500	4 200	1 500	4 200

3.2 Description of variables. Jonsson and Ågren (1992) examined the predictive ability of 32 indices based on the questions in the survey. Here we will only consider the following indices:

AGEBF = the percentage of households that believe that the general economic development in Sweden will improve over the next twelve month

APFBF = the percentage of households that believe that their family's financial situation will improve over the next twelve month

PN6 = the percentage of households that are 100% sure that they will buy a new car within six months

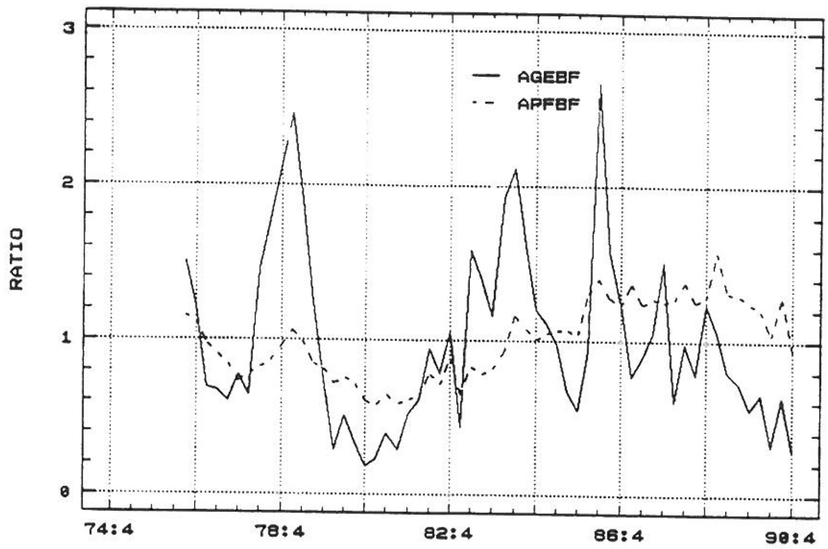
PN12 = the percentage of households that are 100% sure that they will buy a new car within 12 months

PN24 = the percentage of households that are 100% sure that they will buy a new car within 24 months

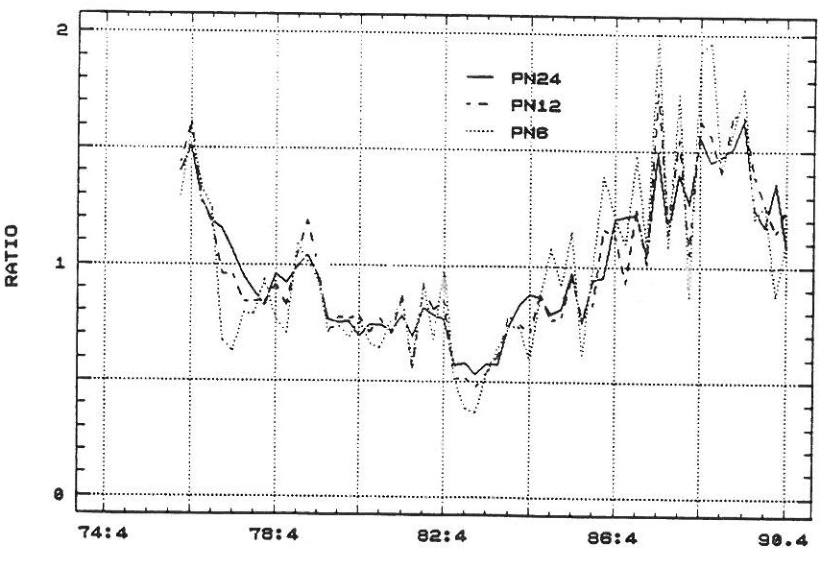
This selection is partly determined by the results in Jonsson and Ågren (1992). The indices are also proportion variables based on individual questions and estimates of the error variances are therefore easily obtained. Since the effect of errors in variables is dependent on the levels of the variables, we wanted the indices to be on different levels and of special interest were indices on low levels, as e.g. the plan variables. The five indices divided by their means are illustrated in Figure 2. For the two attitude indices (AGEBF and APFBF) these ratios show a somewhat different pattern and the most striking feature is the much larger variability for AGEBF. The ratios for the three plan indices (PN6, PN12 and PN24) show a very similar cyclical pattern and a guess is that the corresponding population coefficients of variation are of about the same sizes. Since the means of the three plan variables are rather low (about .8% over the whole period for PN6, 1.6% for PN12 and 3.1% for PN24) we expect the parameters of these variables to be underestimated or heavily underestimated. The

FIGURE 2. The ratios of the attitude and plan indices and their corresponding means during 1976:3 -1990:4.

A: Attitude indices



B. Plan indices



means of the two attitude variables lie on much higher levels, about 12% for AGEBF and 18% for APFBF. Due to the large variability in AGEBF we would expect the OLS bias to be small for that variable, while there might be some effect on the other attitude variable. The plan indices show a pronounced upward trend in the first part of the forecast evaluation period and a negative trend during at least the last two years. The trend break seems to occur slightly earlier for the short horizon index. A similar pattern, but not as pronounced, can be noted for APFBF.

3.3 **Error variances.** As mentioned the indices consist of sample estimates of population proportions. Since these proportions constitute the regressor in Models A and B, the precision with which these numbers are estimated is of great interest. The data needed for calculation of error variances have been obtained from Statistics Sweden. Some of the figures can also be found in official publications. Error variances for the two attitude variables are obtained directly, while the standard errors for the plan variables have been calculated as the standard errors for the number of households that with probability one will buy a new car within i months divided by the estimated population size. Because we are dealing with proportions, estimates of the error variances could also be obtained directly from the observed values of the indices if the sample scheme of the survey is assumed to be simple random sampling. A preliminary analysis shows that the error variances of the plan variables with such an assumption are of the same size on average as those obtained from Statistics Sweden. However, this is not the case for the attitude variable APFBF, for which we get an underestimation of the variances of approximately 16%.

The large sample bias of OLS of the slope parameter in a simple linear regression model is determined by the ratio of the mean error variance in observed X and the variance of X. Since the seasonal dummies in Model A are found to explain almost nothing of the variabiliy in the indices, this ratio also reflects the attenuation of β_1 in that model (see Section 2.3). The ratio of the sum of estimated error variances ($\sum_{1}^{T}\hat{\sigma}_{ut}^{2}$) and $S_{XX}=\sum_{1}^{T}(X_t-\overline{X})^2$ has been calculated for each index and each of the 24 estimation periods. These ratios are illustrated in Figure 3. We note ratios close to zero for the attitude variable AGEBF indicating very small OLS bias. The effect for the other attitude variable (APFBF) is also small but may not be negligible. The ratios for the plan variables are much larger and, as expected, largest for the index on the lowest level (PN6).

According to Jonsson (1992b) the performance of a predictor depends on the relation between its error variance of the prediction period and its mean error variance of the estimation period. In Figure 4 we present the estimated error variances of the 24 forecast periods divided by the mean error variances of the corresponding estimation periods. We note that the error variance of a prediction period with a small survey is always considerably larger than the average error variance of the corresponding estimation period. The reason for that is mainly the reduced sample size over time, but also a higher level of a variable results in a larger error variance. The largest discrepancies between error variances of the prediction and estimation periods occur for the first prediction evaluation period (85:1 to 86:4) and the smallest for the last period. This has to do with increasing mean error variances of the estimation periods over time.

FIGURE 3. The ratio of the sum of estimated error variances and S_{XX} for five indices and the 24 estimation periods (1: *76:3-84:2*, 2: *76:3-84:3*, ..., 24: *76:3-90:1*).

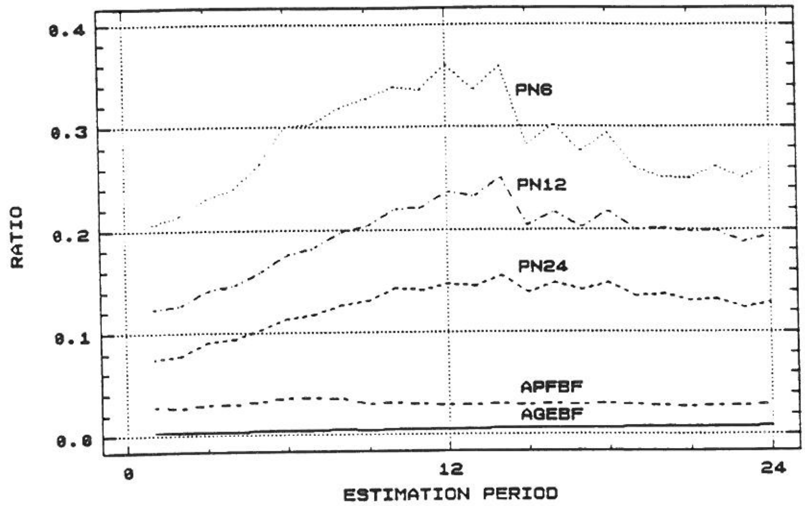
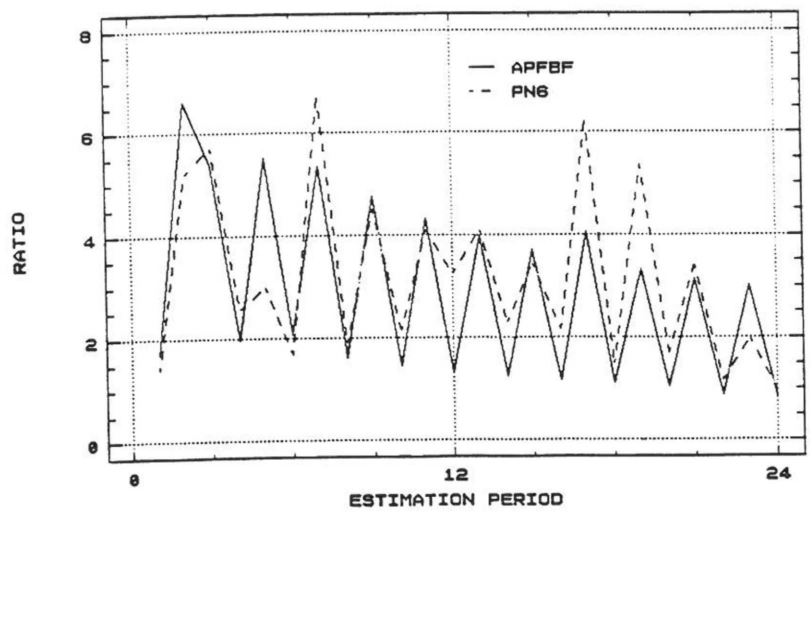


FIGURE 4. The ratios of the estimated error variances of the 24 prediction periods and the mean error variances of the corresponding estimation periods for APFBF and PN6.



4. THE PREDICTORS

The predictors to be used here are based on

a) OLS estimation of the parameters α_i and β_i in (2),

b) consistent estimation of these parameters, and

c) a modified OLS estimator assuming that the error variance of X_t for the estimation period is on average equal to the error variance of the prediction period.

The consistent estimator to be used is proposed by Fuller (1987, pp. 193) and is

$$\mathbf{b} = \hat{\mathbf{H}}_{\mathbf{xx}}^{-1}\hat{\mathbf{H}}_{xy} \qquad (5)$$

where

$$\hat{\mathbf{H}}_{zz} = \sum_{t=1}^{T}\left[\mathbf{Z}_t'\mathbf{Z}_t - (1-\delta T^{-1})\Sigma_{ut}\right] \qquad \text{if } \lambda > 1+T^{-1}$$

and

$$\hat{\mathbf{H}}_{zz} = \sum_{t=1}^{T}\left[\mathbf{Z}_t'\mathbf{Z}_t - (\lambda - T^{-1} - \delta T^{-1})\Sigma_{ut}\right] \qquad \text{if } \lambda \leq 1+T^{-1}$$

and where $Z_t=(Y_t,X_t)$. In Model A, X_t is a vector of ones, seasonal dummies and X_t. Σ_{ut} is a null matrix except for the diagonal element equal to the error variance (σ^2_{ut}) of the variable X_t according to (2). Here λ is the ratio of the residual sum of squares when taking the regression of X_t on the remaining variables in the model (including the dependent variable) and the sum of error variances in X_t. λ is expected to be greater than one, otherwise a modification is made in the estimator. δ is a positive constant that will be set to two in this study. The error variance σ^2_{ut} is estimated according to Section 3.3. The estimator (5) was one of the evaluated consistent estimators in Jonsson (1992a), see Section 1. Because of unequal error variances of the measurements on the regressor a weighted estimator was also considered. Here, this estimator will give higher weights to the oldest than to the most recent observations. Because of that and that a potential gain with a weighted estimator mainly refers to its variance this estimator will not be considered.

To obtain an estimator corresponding to the one based on OLS and a mean error variance in the estimation period of the same size as that of the prediction period, σ^2_{up}, we adjust the estimator (5) by replacing σ^2_{ut} in Σ_{ut} with σ^2_{ut}-σ^2_{up}. This estimator will then have the same probability limits as those obtained by OLS if the mean error variance of the estimation period had been the same as the error variance of the prediction period. The properties of a predictor based on such a modified OLS estimator is discussed in Jonsson (1992b). It is found that this predictor behaves as the corresponding OLS predictor in the case of equal error variances.

The effect of the panel design of the surveys on the estimators should be small in large samples but may influence the results in small samples. To illustrate this, assume a simple relationship and constant error variances. Then the expectation of S_{XX} is

$$E\left[\sum(X_t-\bar{X})^2\right]=E\left[\sum(x_t-\bar{x})^2\right]+E\left[\sum(u_t-\bar{u})^2\right] \qquad (6)$$

The last term in (6) is equal to

$$T\sigma_u^2-E(\Sigma u_t)^2/T \qquad (7)$$

If the errors can be assumed to be independent the expression in (7) is equal to (T-1) times the error variance, which may lead to a choice of δ in (5) equal to one. But here the errors are dependent and if we assume a design where one fifth of the sample is replaced at each survey occasion we get

$$\frac{E(\sum u_t)^2}{T} = \sigma_u^2\left\{1+\frac{2}{T}[(T-1)0.8r_1+(T-2)0.6r_2+(T-3)0.4r_3+(T-4)0.2r_4]\right\}, \quad (8)$$

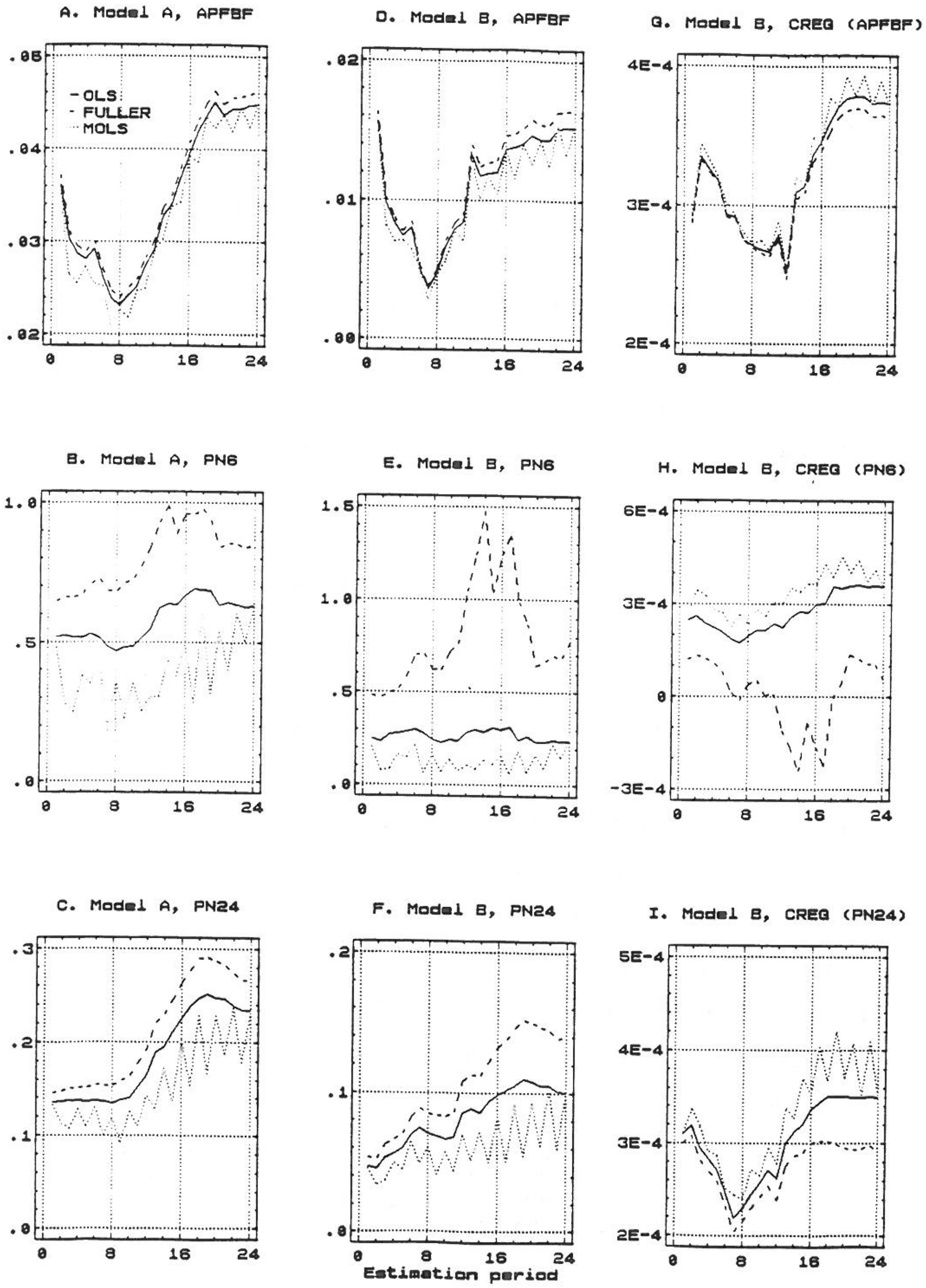
where $r_i = E(u_t u_{t-i})/E(u^2_t)$ for a given panel. It is obvious that if T is large we need not worry about correlated errors. Here T varies from 32 to 55 so we may have to adjust for such dependencies. A small correlation study based on the October survey of 1990, the April survey 1991 and an assumption of that r_i is half the size of r_{i-1} (i>1) could motivate a choice of δ=2 instead of δ=1. However, the exact choice of δ is not critical for the results. Because of a change in panel design and since the estimation period ends three quarters before the prediction quarter, the error in the prediction period will only be correlated with the last error in the estimation period. Furthermore, that correlation can be expected to be very low and we can certainly assume the errors of the prediction periods to be uncorrelated with the errors of the estimation periods.

5. THE RESULTS

5.1 Estimation. Before going into the forecast evaluation let us briefly study the parameter estimates, since we will get different forecasts only if the estimates differ between the predictors. The λ correction in (5) was made only for the plan variable on the lowest level (PN6) and then only for one estimation period. Furthermore the modification made was very small so the sequences of observations are not "bad" according to the λ-criteria. When the attitude variable in the models is AGEBF, the estimators produce very similar estimates, what means that the estimated models will yield the same forecasts. We will therefore not consider that index in the prediction evaluation. The estimates of β_1 (attitude/plan variable) and β_2 (*CREG*="car registrations") in Model A and Model B, when APFBF, PN6 and PN24 are used as attitude/plan index, are illustrated in Figure 5. The consistent estimator will be called the Fuller estimator and the Modified OLS estimator is denoted MOLS. Starting with the results for Model A and the attitude variable APFBF, we note small differences between estimates of different predictors. But, when the prediction periods coincide with small sized surveys (see Section 3.3), the estimates of the MOLS estimator are somewhat smaller than the corresponding OLS estimates. For the plan variables and especially for PN6 we note large differences between the estimators, as expected. The OLS estimates are smaller or much smaller than the Fuller estimates and as mentioned in Section 3.3 the differences between these two estimators should be of the same size if the dummies were excluded from the model. The estimates of the MOLS estimator are overall smaller than those of OLS and much smaller when the prediction error variances are based on small surveys. Noteworthy is also the sometimes large differences in the parameter estimates between different estimation periods. The patterns of the estimates of β_1 in Model 2 follow those of β_1 in Model 1, but the OLS estimates are now smaller and the differences between the OLS and the Fuller estimator are often larger. This depends on a positive correlation between *CREG* and the attitude/plan indices (see Section 2.3). The relation between the estimates of the estimators of β_2 is the opposite to those of β_1. A potential underestimation of β_1 with OLS will therefore at least somewhat be compensated for by an overestimation of β_2. The extreme behaviour of the Fuller estimates for PN6 and *CREG* in Model 2 might be caused by multicollinearity.

5.2 Forecasts using Model A. In this section the forecast evaluation for Model A will be presented. Mean errors (ME) and root mean square errors (RMSE) are shown in Table A in the appendix for one attitude variable (APFBF) and two of the plan

FIGURE 5. Estimates of β_1 and β_2 in model A and B for the 24 estimation periods using OLS, the Fuller estimator and the MOLS estimator.



variables (PN6 and PN24)[2]. The results are presented for the whole forecast period, the period of increase in car expenditures, the period of high level expenditures and the decline period according to Section (2.2). These results are also given separately for small (1500 households) and large surveys (4200 households). This is done because the error variance of an index based on a small survey is much larger than if it is based on a large survey. Hence the results of the prediction evaluation may differ between periods with small and large surveys. However, it should be observed that the number of predictions in the three sub periods then becomes very small (4). The choice of RMSE as a measure of prediction accuracy is motivated from the results in Jonsson(1992b), which are given in terms of MSE. Yet, mean absolute deviations have also been calculated as in Jonsson and Ågren (1992) but will not be discussed further since the results are similar to those for RMSE. At the end of the table we find the mean car expenditures (CA) for each period.

Starting with the attitude index, APFBF, we note that the results are very similar for the three predictors. A certain discrepancy is found for small surveys, where the MOLS predictor seems to be best in the last prediction period, while it is the worst predictor in the second period. This is according to Jonsson (1992b). In the case of greater prediction than estimation error variance he shows for a simple linear regression model that the MOLS predictor is better than the OLS predictor when the true prediction value x_p lies relatively close to the estimation period mean of the regressor, while the OLS predictor is the best one when x_p lies far away from this mean. Looking at the plan variables PN6, PN12 and PN24, an overall evaluation shows the lowest RMSE for OLS except for PN24, where the Fuller predictor is somewhat better. The pattern of a relatively good performance of the MOLS predictor in the third period is also present for the plan variables, but is now much

[2] Tables for Model A with X=PN12 and tables for Model B with X=PN6 and with X=PN24 are to be found in Jonsson, B. Forecasting Car Expenditures Using Household Survey Data - A Comparison of Different Predictors. *Research Report 1993:1, Department of Statistics, Uppsala University, Uppsala.* The report can be ordered from the author.

more pronounced. The Fuller predictor should produce greater RMSE than both the OLS and the MOLS predictor for the last period and that is also the case. Noteworthy is also that the Fuller predictor, with the exception of PN6, is always the best predictor during the second period when the true x values are expected to lie far away from the means of the estimation periods (see Figure 2). For PN6, the OLS and the Fuller predictor do about equally well despite a much larger ME for OLS. It should also be mentioned that looking over the whole period the forecast errors are fairly large for all predictors.

5.3 **Forecasts using Model B.** For Model B we will only give a short summary of the results. All predictors give very similar results when X=APFBF. The results for PN12 are similar to those of PN24. For PN6 the OLS predictor shows a better overall performance than the other two predictors and especially compared to the Fuller predictor with a strikingly bad performance for small surveys in the last evaluation period. The differences between the OLS predictor and the MOLS predictor are not as large as for Model A, but also here the MOLS predictor is often best for evaluation period three and OLS is mostly so for period one and two.

5.4 **Forecasts using smoothed indices.** In Jonsson (1992b) it was shown that the term $B=\beta^{*2}\sigma^2_{up}$, where β^* is the plim value of the estimator of β_1, had a large influence on the variance of the forecast error. If the changes in the true x from one observation to the next could be assumed to be small, one way to reduce the forecast errors would be to use a smoothed value of the index when making the prediction. We will here briefly study this alternative by smoothing the index values during the prediction period according to

$$X_p^S = \frac{N_p}{N_p + N_{p-1}} X_p + \frac{N_{p-1}}{N_p + N_{p-1}} X_{p-1}. \qquad (9)$$

In (9) N_p is the sample size in period p adjusted for non response. If $x_p=x_{p-1}$, then the chosen weights minimize the variance of (9). σ^2_{up} in the MOLS estimator is now replaced with the error variance of (9). The results for Model A with X=APFBF, PN6 or PN24 are given in Table B in the appendix (see also note 2 p. 18). Compared to earlier results for Model A, RMSE now is for all indices lower for the small surveys, while the opposite holds for the large surveys. Better overall RMSE is obtained mainly for PN6 and then especially for the Fuller predictor, but still the OLS predictor is the best one. Looking at the different prediction periods, the gain for the Fuller predictor lies primarily in the second period, where it now is the best predictor. Still it produces relatively bad forecasts for the third period, especially in the case of small surveys. The above pattern of smaller RMSE for the small surveys is also obtained for Model B. However, when the Fuller predictor is used with PN6, here too a large reduction in RMSE can be noted for the large surveys in the second period. The results of this simple way of smoothing are promising and we suggest further studies of how to obtain better estimates of the latest population proportion to reduce the forecast errors.

5. SUMMARY

Recently Jonsson and Ågren (1992) studied the possibility of predicting the Swedish households' expenditures on cars using quick indicators based on survey data concerning households' attitudes towards the economic development and households buying plans. They used OLS when estimating the regression models, but since the data contain sampling errors, other ways of estimating the prediction models may be of interest. Here the study of Jonsson and Ågren has been repeated with a predictor based on consistent estimation and a predictor based on a Modified OLS estimator taking into account different error variances between the estimation and prediction periods. Two models have been used. The first contains an attitude/plan variable and seasonal dummies as regressors. The second model includes additionally a variable measuring car registrations. The attitude and plan variables are proportion variables, some of which are on very low levels.

In the forecast evaluation large differences in RMSE between predictors based on different estimation methods are obtained for the plan variables only, and especially for the one on the lowest level. The performance of the different predictors depends on for which value in an index's cycle we make the prediction and the relationship between the index's error variance of the prediction period and its mean error variance in the estimation period. To base the predictor on OLS is often better than to base it on consistent estimation. The MOLS predictor turns out to be better than the OLS predictor when an index's prediction value lies close to its mean. It is also worth mentioning that in the case of large error variances it can be worth trying to smooth the index to obtain better predictions. The results coincide on the whole with those of Jonsson (1992b). A special feature of the data used was that the error variance of the prediction period often was larger than the mean error variance of the estimation period. It is important to emphasise that if the error variance of the prediction period had been the smallest, then the Modified OLS predictor and the Fuller predictor probably had been more competitive to the one based on OLS.

REFERENCES

Ågren, A. (1989), A Survey of some Work on the Predictive Value of Attitude Data in Consumption and Saving Models. *Research Report 1989:2, Department of Statistics*, Uppsala University, Uppsala.

Bartlett, M. S. (1951), An Inverse Matrix Adjustment Arising in Discriminant Analysis, *Annals of Mathematical Statistics*, vol. 22, pp.107-11

Chow, Gregory C. (1957), *Demand for Automobiles in the United States*, Amsterdam: North-Holland.

Fuller, W.A. (1987), *Measurement Error Models*, New York: John Wiley & Sons.

Jonsson, B. (1992a), Sample Based Proportions as Values on an Independent Variable in a Regression Model. National Bureau of Economic Research (Sweden) Working paper 16, 1992.

Jonsson, B. (1992b), Prediction with a Linear Regression Model and Errors in a Regressor. National Bureau of Economic Research (Sweden) Working paper 19, 1992.

Jonsson, B. and Ågren A. (1992), Forecasting Car Expenditures Using Household Survey Data. National Bureau of Economic Research (Sweden) Working paper 9, 1992.

Kamakura, W.A. and Gessner G. (1986), Consumer Sentiment and Buying Intentions Revisited: A Comparison of Predictive Usefulness. *Journal of Economic Psychology* 7, 197-220.

Levi, M.D. (1973), Errors in the Variables Bias in the Presence of Correctly Measured Variables. *Econometrica* **41**, 985-6.

Pickering, J.F. (1984), Purchase Expectations and the Demand for Durables. *Journal of Economic psychology* 5, 341-52.

APPENDIX. TABLE 1A. Results of the forecast evaluation for Model A.

Period Index	85:1-90:4 ME	85:1-90:4 RMSE	85:1-86:4 ME	85:1-86:4 RMSE	87:1-88:4 ME	87:1-88:4 RMSE	89:1-90:4 ME	89:1-90:4 RMSE
APFBF all surveys								
OLS	439	1028	-39	543	1469	1592	-113	587
FULLER	417	1018	-58	542	1446	1570	-138	592
MOLS	478	1036	3	540	1502	1615	-71	564
APFBF small surveys								
OLS	204	783	-108	428	1053	1143	-333	592
FULLER	180	781	-125	434	1032	1124	-368	616
MOLS	274	796	-40	418	1111	1196	-251	545
APFBF large surveys								
OLS	674	1225	30	637	1886	1939	107	582
FULLER	654	1210	10	632	1860	1915	91	566
MOLS	682	1230	46	640	1893	1946	108	583
PN6 all surveys								
OLS	321	946	208	651	1110	1321	-355	721
FULLER	-138	1156	91	637	363	1324	-869	1361
MOLS	649	1064	321	682	1544	1635	83	512
PN6 small surveys								
OLS	342	1011	141	459	1395	1472	-509	828
FULLER	28	1318	36	498	1229	1327	-1182	1789
MOLS	703	1008	299	521	1556	1624	255	378
PN6 large surveys								
OLS	300	877	274	798	825	1149	-201	594
FULLER	-304	969	147	751	-504	1321	-555	710
MOLS	596	1118	344	811	1533	1646	-89	617
PN24 all surveys								
OLS	407	1012	423	678	1356	1469	-560	672
FULLER	269	998	412	663	1192	1335	-798	873
MOLS	551	1062	428	689	1518	1612	-292	559
PN24 small surveys								
OLS	355	902	427	597	1193	1260	-554	704
FULLER	246	921	430	597	1104	1174	-796	899
MOLS	561	949	422	598	1370	1434	-110	537
PN24 large surveys								
OLS	458	1111	419	751	1520	1652	-565	638
FULLER	292	1069	395	724	1280	1479	-800	846
MOLS	542	1164	434	770	1666	1771	-475	579
All surveys								
Mean CA	4262		3459		5303		4025	

TABLE B. Results of the forecast evaluation in the smoothed case. Model A.

Period Index	85:1-90:4 ME	85:1-90:4 RMSE	85:1-86:4 ME	85:1-86:4 RMSE	87:1-88:4 ME	87:1-88:4 RMSE	89:1-90:4 ME	89:1-90:4 RMSE
APFBF all surveys								
OLS	471	1008	-2	529	1476	1588	-60	495
FULLER	450	995	-20	526	1452	1567	-83	488
MOLS	472	1008	6	528	1475	1588	-64	495
APFBF small surveys								
OLS	322	720	-42	397	1059	1136	-52	329
FULLER	301	709	-57	398	1038	1117	-77	324
MOLS	324	720	-33	391	1059	1137	-56	329
APFBF large surveys								
OLS	621	1230	38	635	1893	1938	-69	618
FULLER	598	1215	18	629	1867	1914	-90	609
MOLS	621	1230	45	636	1892	1937	-72	617
PN6 all surveys								
OLS	238	850	239	655	973	1118	-497	699
FULLER	-250	975	139	616	184	774	-1073	1369
MOLS	365	910	272	658	1220	1330	-396	571
PN6 small surveys								
OLS	-3	687	143	475	617	636	-768	887
FULLER	-515	1141	52	455	-61	590	-1535	1831
MOLS	135	683	176	467	841	854	-613	672
PN6 large surveys								
OLS	479	986	334	795	1330	1447	-227	438
FULLER	15	774	226	744	429	922	-611	629
MOLS	596	1104	367	805	1599	1676	-178	446
PN24 all surveys								
OLS	356	1012	420	679	1322	1437	-674	741
FULLER	209	1009	410	664	1153	1288	-936	976
MOLS	380	1028	422	682	1365	1476	-646	723
PN24 small surveys								
OLS	189	785	367	534	951	971	-753	786
FULLER	45	836	364	527	812	825	-1041	1068
MOLS	213	795	366	535	990	1014	-718	761
PN24 large surveys								
OLS	524	1197	473	798	1694	1785	-595	692
FULLER	374	1155	457	777	1495	1624	-831	874
MOLS	547	1217	477	803	1740	1825	-575	684

Sammanfattning

Jonsson och Ågren (1992) studerade nyligen vilket värde hushållens attityder om den ekonomiska utvecklingen och deras planer avseende bilinköp har då det gäller att prognostisera de svenska hushållens totala utgifter på bilar. På basis av SCB's så kallade HIP undersökningar bildades flera olika index som sedan användes som regressorer för att förklara och predicera bilutgifterna. Jonsson och Ågren skattade sina regressionsmodeller på traditionellt sätt med vanlig minsta-kvadrat metod (OLS), men eftersom indexen innehåller samplingfel kan andra estimatorer vara värda att pröva. Jonsson och Ågrens studie har här upprepats med en prediktor baserad på konsistent skattning av modellens parametrar och en modifierad OLS-estimator som tar hänsyn till eventuella olikheter mellan mätfelsvarianserna för estimations- och prediktions perioderna. Två modeller har använts. Den första har som regressorer en attityd/plan variabel och säsongdummies. Den andra inkluderar också en variabel baserad på bilregistreringsstatistik. Attityd och planvariablerna utgörs av andelar som för planvariablerna är mycket låga.

Endast för planvariablerna visade prognosutvärderingen stora skillnader i resultaten för de olika estimationsmetoderna. I synnerhet var detta fallet för planvariabeln på den lägsta nivån. Hur pass bra de olika prediktorerna är i förhållande till varandra beror på det värde i indexvariabelns cykel för vilket prognosen görs och relationen mellan variabelns felvarians i estimations- och prediktionsperioden. Överlag verkar det vara bättre att använda sig av en OLS baserad prediktor än en prediktor baserad på konsistent skattningsmetodik. Den modifierade OLS-prediktorn visar sig vara bättre än OLS- prediktorn när värdet på det index, för vilket man vill ha en prognos, ligger nära medelvärdet för estimationsperioden. Det är också värt att nämna att för fall med stora mätfelsvarianser kan det vara värt att försöka med utjämnade index för att erhålla bättre prognoser. De empiriska resultaten överensstämmer i stort med de teoretiska i Jonsson (1992b). En speciell egenskap hos de använda indexserierna är att felvarianserna är större under prediktionsperioderna än under estimationsperioderna. Om det motsatta hade varit fallet hade förmodligen prognoser baserade på den modifierade estimatorn och den konsistenta estimatorn varit mer konkurrenskraftiga i förhållande till OLS.

Prevoius titles in this serie:

No. 1	Current Account and Business Cycles: Stylized Facts for Sweden by Anders Warne and Anders Vredin. December 1989.
No. 2	Change in Technical Structure of the Swedish Economy by Göran Östblom. December 1989.
No. 3	Mamtax. A Dynamic CGE Model for Tax Reform Simulations by Paul Söderlind. December 1989.
No. 4	The Supply Side of the Econometric Model of the NIER by Alfred Kanis and Aleksander Markowski. November 1990.
No. 5	The Financial Sector in the SNEPQ Model by Lennart Berg. February 1991.
No. 6	Consumer Attitudes, Buying Intentions and Consumption Expenditures. An Analysis of the Swedish Household Survey Data by Anders Ågren & Bo Jonsson. April 1991.
No. 7	A Quarterly Consumption Funtion for Sweden 1979-1989 by Lennart Berg & Reinhold Bergström. October 1991.
No. 8	Good Business Cycle Forecasts - A Must for Stabilization Policies by Lars-Erik Öller. February 1992.
No. 9	Forecasting Car Expenditures Using Household Survey Data by Bo Jonsson and Anders Ågren. February 1992.
No. 10	Forecasting the Business Cycle Not Using Minimum Autocorrelation Factors by Karl-Gustaf Löfgren, Bo Ranneby & Sara Sjöstedt. February 1992.
No. 11	Current Quarter Forecasts of Swedish GNP Using Monthly Variables by Stefan Gerlach. February 1992.
No. 12	The Relationship Between Manufacturing Production and Different Business Survey Series in Sweden by Reinhold Bergström. February 1992.
No. 13	Forecasting the Swedish Unemployment Rate: V A R vs Transfer Function Modelling by Per-Olov Edlund and Sune Karlsson. March 1992.
No. 14	Business Survey Data in Forecasting the Output of Swedish and Finnish Metal and Engineering Industries: A Kalman Filter Approach by Markku Rahiala and Timo Teräsvirta. March 1992.

No. 15 The Relationship Between Manufacturing and Various BTS Series in Sweden Illuminated By Frequency and Complex Demodulate Methods by Anders Christoffersson, Roland Roberts and Ulla Eriksson. April 1992.

No. 16 Sample Based Proportions As Values On An Independent Variable In A Regression Model by Bo Jonsson. October 1992.

No. 17 Eliciting Turning Point Warnings From Business Surveys by Lars-Erik Öller. November 1992.

No. 18 Volatility, Trading Mechanisms and International Cross-Listing by Margaret M. Forster. December 1992.

No. 19 Prediction With a Linear Regression Model and Errors in a Regressor by Bo Jonsson. December 1992.

No. 20 Corporate Control, Portfolio Choice, and the Decline of Banking by Gary Gorton and Richard Rosen. January 1993.

No. 21 The Index of Industrial Production. A Formal Description of the Process Behind It by Claes-Håkan Gustafsson and Åke Holmén. February 1993.

No. 22 A General Equilibrium Analysis of the Swedish Tax Reforms 1989-91 by Tohmas Karlsson. February 1993.