Enache, Daniel; Weihs, Claus

Working Paper
Importance Assessment of Correlated Predictors in Business Cycles Classification

Technical Report, No. 2004,66

Provided in Cooperation with:
Collaborative Research Center 'Reduction of Complexity in Multivariate Data Structures' (SFB 475), University of Dortmund


This Version is available at:
http://hdl.handle.net/10419/22579

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
Importance Assessment of Correlated Predictors in Business Cycles Classification

Daniel Enache and Claus Weihs

University of Dortmund*
Department of Statistics
44221 Dortmund, Germany

Abstract. When trying to interpret estimated parameters the researcher is interested in the (relative) importance of the individual predictors. However, if the predictors are highly correlated, the interpretation of coefficients, e.g. as economic “multipliers”, is not applicable in standard regression or classification models. The goal of this paper is to develop a procedure to obtain such measures of importance for classification methods and to apply them to models for the classification of German business cycle phases.

1 Problem

1.1 Introduction

Multivariate classification of the four business cycle phases upswing, upper turning point, downswing, and lower turning point is often performed by linear discriminant analysis (LDA, cf. Meyer and Weinberg (1975)) and by time series analysis methods (e.g. Krolzig (1997)). Lately, other classification methods, like quadratic discriminant analysis, classification trees, artificial neural networks and support vector machines, have also been applied to this problem (e.g. Garczarek and Weihs (2002)) and new classification methods have been developed to solve this problem (e.g. Röhl et al. (2002)).

Heilemann and Münch (1996) reduced the stylized facts to a set of 13 important variables (see also Theis et al. (1999), Weihs and Garczarek (2002)):

* This work has been supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 475.
In analyzing business cycles it is important not only to obtain good predictions, but also to measure the influence of the individual variables to get an impression of their importance. In linear models, the measures of influence are usually the regression weights. Under ceteris-paribus assumptions, these weights measure how much the dependent variable changes if the independent variable is varied by a certain amount. In economic contexts these regression coefficients are called “multipliers”.

Similar to these regression models, the coefficients of linear classification equations, like LDA, can be interpreted as influences on the probability of being classified into the selected class. For the interpretation of the coefficients in linear regression or classification models it is crucial, that the independent variables are uncorrelated. Unfortunately, several “stylized facts” appear to be highly correlated, which prevents the interpretation of the coefficients as “multipliers”, since the ceteris-paribus assumption is not realistic.

1.2 Measures of Importance

This paper focuses on the importance of individual variables on the classification of business cycle phases. According to the focus of an analysis, two types of statistical importance should be distinguished: the importance with respect to model selection and the importance with respect to value change.

The first type is based on all the measures used for model selection, like F-values (e.g. F-to-enter, F-to-remove), $R^2$, etc. (e.g. Rencher(1995)). The second type is based on the measures specifying value changes of the dependent variable if the predictor variable changes its value. For regression and linear classification models these are usually the coefficients to be estimated.

2 Correlated Predictors in Regression Models

2.1 Overview

In order to develop an approach for measuring the importance of correlated predictors in classification models, it can be useful to discuss some results in
regression models, because a lot of research concerning correlated predictors has been done in this area. The assumption of uncorrelated predictors is often not appropriate for macroeconomic data. In fact, some of the variables are highly correlated. This usually leads to correlated regression coefficients, which are not easily interpreted (Assenmacher (2002)) and the coefficients cannot be interpreted as “multipliers”.

In a regression model containing correlated predictor variables, there exist several approaches to handle highly correlated variables:

The first type of methods transforms the predictor variables to eliminate the correlations. For example, orthogonalization of predictors, which is often carried out by sequential regression (e.g. Kruskal (1987)). Such methods allow the interpretation of the variables. On the other hand, the coefficients highly depend on the order of variables entered into the model.

The second type of methods corrects the coefficients using a scalar shrinkage parameter, like ridge regression (e.g. Hoerl and Kennard (1969)). The drawback of these methods is that the scalar added to the main diagonal of the covariance matrix does not improve interpretability.

The third type of methods tries to collect correlated variables into latent variable, thus reducing dimensionality. These methods often use principal component regression models (e.g. Hawkins (1973)). The principal components representation does also not allow the interpretation of the single variable effect under ceteris paribus conditions. Models which combine ridge regression and principal component regression have also been presented (e.g. Stone and Brooks (1990)), but have the same drawbacks as the individual approaches.

2.2 Orthogonalization

Because the focus of this paper is on the interpretation of the single variable’s influences, an orthogonalization method which is based upon sequential regression has been used here to address the multicollinearity problem. The disadvantage of this approach is, that the coefficients highly depend on the order, in which the variables have been entered into the model.

Based upon Kruskal’s (1987) idea, Fickel (2002) proposes an algorithm to overcome this disadvantage. It estimates sequential regression models for every sequence \(i\) out of the \(p!\) possible variable sequences. From these estimations the coefficients \(\beta_{ij}\) and the increase in the coefficient of determination \((\Delta R^2)_{ij}\) are stored for each variable \(j\) and each sequence \(i\). Then for each variable

\[
\hat{\gamma}_j = \frac{1}{p!} \sum_{i=1}^{p!} \hat{\beta}_{ij}, \quad \delta_j = \frac{1}{p!} \sum_{i=1}^{p!} (\Delta R^2)_{ij}
\]

(1)

are estimated. The average coefficient \(\hat{\gamma}_j\) is a measure of importance for value change and can be interpreted as average effect of variable \(j\), when all other...
variables are held constant (like “multipliers”) and the average $R^2$-increment $\hat{\delta}_j$ as relative importance of variable $j$ for model selection.

The additional level shift of the residuals $e_2, \ldots, e_p$ introduced in Fickel’s paper in order to scale the residuals to the same level as the original variables is not used here, since the business cycle data consist of growth rates.

3 Correlated Predictors in Classification Models

3.1 Orthogonalization

Most classification models provide a method to estimate the membership probability of each class $k$, $k = 1, \ldots, K$, $p_{Mod}(k|X)$, using specific density assumptions and specific estimation criteria.

The approach of this paper is to use a linear probability model to estimate the importance of the correlated variables for the estimated class membership probabilities:

$$\hat{p}_{Mod}(k|X) = \alpha_k + X\beta_k + \epsilon_k, \quad k = 1, \ldots, K.$$  \hspace{1cm} (2)

where $X$ is the matrix containing a sample of the random variable $x$. This is done by estimation of the $K$ discriminant functions by an appropriate classification model $Mod$ and of the posterior probabilities $\hat{p}_{Mod}(k|x_i)$. These are then used as dependent variables in $K$ linear regression models, one for each class.

Fickel’s (2002) method can now be applied to each of the individual regression equations (2). For each class $k = 1, \ldots, K$, all $p!$ possible variable sequences are used to estimate sequential regression models. Then the importance measures $\hat{\gamma}$ and $\hat{\delta}$ are computed from the estimation results.

This orthogonalization procedure can be used for a great variety of classification methods. The only requirement is that the classification is based upon a membership function $m_{Mod}(k|x)$ or, even better, upon the estimated posterior probability function $\hat{p}_{Mod}(k|x)$. The usage of the posterior probability instead of the membership function $m(k|x)$ enables the comparability of the results of different classification methods. In this paper the results of a linear discriminant analysis and a multinomial logit are compared.

3.2 Using a large Number of Variables

The computation time of the proposed method increases excessively with the number of variables. The reason is that all possible $p!$ permutations of variable sequences have to be evaluated. One possible way to deal with this problem and to obtain interesting results is to choose a random subset of the $p!$ variable sequences.

The chosen subset must be uniformly distributed among the permutations of variable sequences. If the sample of variable sequences is big enough, the means of the coefficients and $R^2$-increments will be estimated well enough.
3.3 Results for the Business Cycle Model

Using a random selection of variable orderings, all 13 “stylized facts” can now be used for the analysis. Instead of evaluating all $13! = 6,227,020,800$ possible permutations of variable sequences, which takes a very long time even on fast computers, only 50,000 randomly chosen variable sequences are used for the analysis. The estimated total correlation matrix for all 13 stylized facts is shown in Table 1. A few variable pairs, like PY and M1 as well as C and PY are almost not correlated, but for most variable pairs, correlations are in effect. The highest positive correlations have LC and PY with 0.868 and C and Y with 0.776. RL and LC (-0.656) and RL and PY (-0.656) have the highest negative correlations. The variables GD, X and M1 do not have very high correlations with other variables.

Please note that bivariate correlations give only a vague impression of the underlying multicollinearity, which should be reflected in the corrections made by the orthogonalization procedure. Table 2 shows the estimated coefficients for the upswing class. The first column $\hat{\beta}_j$ contains the standard regression coefficients for the comparison to the importance measure $\gamma_j$ obtained by the orthogonalization procedure. The most important variables for model selection are RS, C, PC, PY, LC, and IE. PY and RL have been corrected strongly in LDA and in the logit model. Table 3 shows the estimated coefficients for the upper turning point class. The most important variables are C, L, and Y. The greatest correction has been made for Y in LDA and for X in the logit model. RL and LC have also been corrected substantially in both models. Table 4 shows the estimated coefficients for the downswing class. The most important variables for model selection are RS, LC, PY, IE, and RL. The most substantial corrections can be observed for PY and RL in both models. Table 5 shows the estimated coefficients for the lower turning point class. The variables most important for model selection are L and Y. The strongest corrections are observed for RL and for PY in both models.
4 Discussion and Outlook

An orthogonalization procedure has been proposed for classification models with correlated predictor variables. The procedure has been applied to west german business cycle data to model the four cycle phases upswing, upper turning point, downswing, and lower turning point. For 13 pre-selected stylized facts the classification models linear discriminant analysis and multinomial logit have been compared.
For model selection, RS, PY, LC, and IE seem to be important for both the upswing and downswing phases, whereas Y and L seem to have more importance for the turning point phases. For the upswing class additionally the consumption related variables C and PC seem to be important for model selection (C even for the upper turning point class) and RL seems to be characteristic for the downswing class.

The orthogonalization procedure corrects the coefficients for the predictors in such a way, that these corrected coefficients can be interpreted similar to “multipliers”. The procedure also allows to compare different classification
models, as LDA and multinomial logit for this paper. Apparently, the results for both methods are similar for the business cycle data.

Similar comparisons will be done using other classification methods like Support Vector Machines. Also, the approximation of using a linear probability model should be overcome. These two threads will be followed during further research in this area.

References


