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)):

Y   yearly growth rate of the gross national product (GNP)
C   yearly growth rate of private consumption
GD government deficit as a proportion of the GNP
L   yearly growth rate of the number of wage and salary earners
X   netto exports as a proportion of the GNP
M1 yearly growth rate of money supply
IE   yearly growth rate of equipment investments
IC   yearly growth rate of construction investments
LC yearly growth rate of labour unit costs
PY yearly growth rate of GNP price deflator
PC yearly growth rate of the consumer price index
RS nominal short term interest rate
RL real long term interest rate

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 macro economic 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 (eg. 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 $\hat{\beta}_{ij}$ 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

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

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, $k = 1, \ldots, K$, $p_{Mod}(k|\boldsymbol{X})$, 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:

$$\hat{p}_{Mod}(k|\boldsymbol{X}) = \alpha_k + \boldsymbol{X}\boldsymbol{\beta}_k + \boldsymbol{\varepsilon}_k, \quad k = 1, \ldots, K. \tag{2}$$

where $\boldsymbol{X}$ is the matrix containing a sample of the random variable $\boldsymbol{x}$. This is done by estimation of the K discriminant functions by an appropriate classification model Mod and of the posterior probabilities $\hat{p}_{Mod}(k|\boldsymbol{x}_i)$. 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 $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\delta}}$ 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 $m_{Mod}(k|\boldsymbol{x})$ or, even better, upon the estimated posterior probability function $\hat{p}_{Mod}(k|\boldsymbol{x})$. The usage of the posterior probability instead of the membership function $m(k|\boldsymbol{x})$ 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

	Y	C	GD	L	X	M1	IE	IC	LC	PY	PC	RS	RL
Y	1.000												
C	0.776	1.000											
GD	0.403	0.387	1.000										
L	0.737	0.657	0.365	1.000									
X	-0.123	-0.154	-0.195	-0.074	1.000								
M1	0.318	0.423	-0.098	0.198	0.169	1.000							
IE	0.742	0.647	0.257	0.669	-0.160	0.314	1.000						
IC	0.680	0.518	0.279	0.505	-0.042	0.176	0.388	1.000					
LC	-0.170	0.108	0.180	0.153	-0.328	-0.139	-0.087	-0.179	1.000				
PY	-0.176	0.012	0.034	0.048	-0.257	0.000	-0.093	-0.175	0.868	1.000			
PC	-0.352	-0.347	-0.193	-0.203	-0.294	-0.143	-0.367	-0.270	0.567	0.723	1.000		
RS	-0.241	-0.308	-0.181	0.051	0.071	-0.359	-0.269	-0.187	0.426	0.493	0.616	1.000	
RL	-0.094	-0.365	-0.322	-0.226	0.201	-0.197	-0.185	-0.099	-0.656	-0.656	-0.118	0.156	1.000

Table 1. Empirical correlation matrix for the 13 "stylized facts".

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 γ_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.

	LDA			logit		
	$\hat{\beta}_j$	$\hat{\gamma}_j$	$\hat{\delta}_j$	$\hat{\beta}_j$	$\hat{\gamma}_j$	$\hat{\delta}_j$
const.	0.881	0.497	0.000	0.830	0.464	0.000
Y	0.042	0.030	0.031	0.034	0.024	0.023
C	-0.117	-0.081	**0.115**	-0.104	-0.077	**0.093**
GD	-0.025	-0.012	0.007	-0.034	-0.024	0.014
L	0.101	0.058	0.035	0.092	0.046	0.024
X	0.053	0.036	0.034	0.051	0.035	0.029
M1	-0.030	-0.017	0.033	-0.032	-0.017	0.030
IE	0.012	0.016	**0.073**	0.013	0.016	**0.065**
IC	0.000	0.001	0.006	0.001	-0.000	0.004
LC	-0.008	-0.035	**0.076**	-0.022	-0.045	**0.083**
PY	0.162	*-0.006*	**0.075**	0.157	*-0.008*	**0.071**
PC	-0.074	-0.085	**0.102**	-0.054	-0.071	**0.076**
RS	-0.161	-0.105	**0.191**	-0.154	-0.100	**0.156**
RL	0.104	*0.014*	0.025	0.100	*0.023*	0.025

Table 2. Estimated coefficients for the upswing class.

	LDA			logit		
	$\hat{\beta}_j$	$\hat{\gamma}_j$	$\hat{\delta}_j$	$\hat{\beta}_j$	$\hat{\gamma}_j$	$\hat{\delta}_j$
const.	-0.356	0.077	0.000	-0.292	0.093	0.000
Y	-0.017	*0.014*	**0.075**	-0.005	0.023	**0.078**
C	0.056	0.050	**0.149**	0.048	0.046	**0.111**
GD	-0.009	-0.012	0.016	-0.004	-0.006	0.007
L	0.055	0.066	**0.132**	0.043	0.064	**0.109**
X	-0.005	0.003	0.005	-0.015	*-0.068*	0.005
M1	0.011	0.014	0.063	0.011	0.013	0.042
IE	0.000	0.004	0.046	0.003	0.007	0.061
IC	-0.001	0.002	0.024	-0.002	0.002	0.024
LC	-0.029	*-0.014*	0.017	-0.031	*-0.014*	0.013
PY	-0.014	-0.003	0.010	-0.013	0.010	0.009
PC	0.022	0.012	0.012	0.025	0.018	0.013
RS	0.045	0.037	0.069	0.054	0.041	0.062
RL	0.003	*0.026*	0.023	-0.016	*0.013*	0.011

Table 3. Estimated coefficients for the upper turning point class.

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.

	LDA			logit		
	$\hat{\beta}_j$	$\hat{\gamma}_j$	δ_j	$\hat{\beta}_j$	$\hat{\gamma}_j$	δ_j
const.	1.154	0.252	0.000	1.136	0.259	0.000
Y	0.006	-0.013	0.019	0.021	-0.007	0.018
C	0.033	0.022	0.019	0.021	0.015	0.013
GD	0.005	0.015	0.010	0.002	0.015	0.011
L	-0.046	0.002	0.017	-0.017	0.024	0.020
X	-0.051	-0.033	0.035	-0.026	-0.015	0.012
M1	0.008	-0.010	0.036	0.004	-0.013	0.036
IE	-0.015	-0.017	**0.092**	-0.018	-0.019	**0.106**
IC	-0.003	0.000	0.004	-0.004	-0.001	0.004
LC	0.041	0.053	**0.111**	0.064	0.067	**0.130**
PY	-0.336	*-0.094*	**0.093**	-0.348	*-0.108*	**0.090**
PC	0.058	0.043	0.050	0.072	0.047	0.044
RS	0.179	0.110	**0.244**	0.145	0.094	**0.167**
RL	-0.269	*-0.107*	**0.075**	-0.251	*-0.103*	**0.065**

Table 4. Estimated coefficients for the downswing class.

	LDA			logit		
	$\hat{\beta}_j$	$\hat{\gamma}_j$	δ_j	$\hat{\beta}_j$	$\hat{\gamma}_j$	δ_j
const.	-0.679	0.169	0.000	-0.673	0.186	0.000
Y	-0.032	-0.031	**0.070**	-0.050	-0.041	**0.075**
C	0.028	0.009	0.031	0.035	0.016	0.031
GD	0.029	0.010	0.016	0.037	0.015	0.018
L	-0.109	-0.126	**0.264**	-0.119	-0.134	**0.238**
X	0.003	-0.007	0.008	-0.010	-0.013	0.008
M1	0.010	0.013	0.033	0.016	0.016	0.039
IE	0.002	-0.003	0.041	0.002	-0.004	0.038
IC	0.003	-0.004	0.030	0.007	-0.002	0.022
LC	-0.005	-0.004	0.014	-0.012	-0.008	0.012
PY	0.188	*0.103*	0.066	0.203	*0.105*	0.051
PC	-0.006	0.030	0.041	-0.043	0.005	0.020
RS	-0.063	-0.042	0.063	-0.045	-0.034	0.038
RL	0.162	*0.068*	0.043	0.168	*0.069*	0.035

Table 5. Estimated coefficients for the lower turning point class.

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

ASSENMACHER, W. (2002): *Einführung in die Ökonometrie*. 6. Auflage, Oldenbourg Verlag, München.

FICKEL, N. (2002): Regression Analysis of Extremely Multicollinear Data. In: W. Gaul and G. Ritter (Eds.): *Classification, Automation, and New Media*. Springer, Berlin, 67–74.

GARCZAREK, U. M. and WEIHS, C. (2002): Incorporating background knowledge for better prediction of cycle phases. To be published in Knowledge and Information Systems.

HAWKINS, D. M. (1973): On the Investigation of Alternative Regressions by Principal Component Analysis. *Journal of the Royal Statistical Society, Series C: Applied Statistics, 22(1)*, 275–286.

HEILEMANN, U. and MÜNCH, H. J. (1996): West german business cycles 1963–1994: A multivariate discriminant analysis. In: *CIRET-Conference in Singapore, CIRET-Studien 50*.

HOERL, A. E. and KENNARD, R. W. (1969): Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics, 12(1)*, 55–67.

KROLZIG, H.-M. (1997): *Markov-Switching Vector Autoregressions. Modelling, Statistical Inference and Application to Business Cycle Analysis*. Springer, Berlin.

KRUSKAL, W. (1987): Relative importance by averaging over orderings. *The American Statistician, 41*, 6–10.

MEYER, J. R. and WEINBERG, D. H. (1975): On the classification of economic fluctuations. *Explorations in Economic Research, 2*, 167–202.

RENCHER, A. C. (1995): *Methods of Multivariate Analysis*. Wiley, New York.

RÖHL, M. C. WEIHS, C., and THEIS, W. (2002): Direct minimization of error rates in multivariate classification. *Computational Statistics, 17*, 29–46.

STONE, M. and BROOKS, R. J. (1990): Continuum Regression: Cross-validated Sequentially Constructed Prediction Embracing Ordinary Least Squares, Partial Least Squares and Principal Component Analysis. *Journal of the Royal Statistical Society, Series B, 52(2)*, 237–269.

THEIS, W., VOGTLÄNDER, K. and WEIHS, C. (1999): Descriptive Studies on Stylized Facts of the German Business Cycle. Sonderforschungsbereich 475, Technical Report 45/1999, Universität Dortmund.

WEIHS, C. and GARCZAREK, U. (2002): Stability of multivariate representation of business cycles over time. Sonderforschungsbereich 475, Technical Report 20/2002, Universität Dortmund.