Accepted Manuscript

Oracle inequalities, variable selection and uniform inference in high-dimensional correlated random effects panel data models

Anders Bredahl Kock

PII: S0304-4076(16)30122-1
DOI: http://dx.doi.org/10.1016/j.jeconom.2016.06.001
Reference: ECONOM 4279

To appear in: *Journal of Econometrics*

Received date: 17 April 2016
Revised date: 17 April 2016
Accepted date: 19 June 2016

Please cite this article as: Kock, A.B., Oracle inequalities, variable selection and uniform inference in high-dimensional correlated random effects panel data models. *Journal of Econometrics* (2016), http://dx.doi.org/10.1016/j.jeconom.2016.06.001

ORACLE INEQUALITIES, VARIABLE SELECTION AND UNIFORM INFERENCE IN HIGH-DIMENSIONAL CORRELATED RANDOM EFFECTS PANEL DATA MODELS

ANDERS BREDAHL KOCK
AARHUS UNIVERSITY AND CREATES

Abstract. In this paper we study high-dimensional correlated random effects panel data models. Our setting is useful as it allows including time invariant covariates as under random effects yet allows for correlation between covariates and unobserved heterogeneity as under fixed effects. We use the Mundlak-Chamberlain device to model this correlation. Allowing for a flexible correlation structure naturally leads to a high-dimensional model in which least squares estimation easily becomes infeasible with even a moderate number of explanatory variables.

Imposing a combination of sparsity and weak sparsity on the parameters of the model we first establish an oracle inequality for the Lasso. This is valid even when the error terms are heteroskedastic and no structure is imposed on the time series dependence of the error terms.

Next, we provide upper bounds on the sup-norm estimation error of the Lasso. As opposed to the classical ℓ_1- and ℓ_2-bounds the sup-norm bounds do not directly depend on the unknown degree of sparsity and are thus well suited for thresholding the Lasso for variable selection. We provide sufficient conditions under which thresholding results in consistent model selection. Pointwise valid asymptotic inference is established for a post-thresholding estimator. Finally, we show how the Lasso can be desparsified in the correlated random effects setting and how this leads to uniformly valid inference even in the presence of heteroskedasticity and autocorrelated error terms.

Keywords: Panel data, Lasso, oracle inequality, sup-norm bounds, high-dimensional models, weak sparsity, correlated random effects, Mundlak-Chamberlain, variable selection, uniform inference.
JEL-codes: C01, C10, C23.

1. Introduction

In this paper we study panel data models under correlated random effects. As we will see, these models naturally become high-dimensional when the correlation between the covariates and the unobserved heterogeneity is to be modeled in a flexible way. The

I am grateful to Mehmet Caner and Peter Phillips for urging me to pursue the ideas of this paper. I would also like to thank seminar participants at PUC-Rio and participants at the European Meeting of Statisticians 2013 in Budapest for helpful comments and suggestions. The paper has also benefitted tremendously from excellent comments by two anonymous referees as well as the associate and co-editor Jianqing Fan. Financial support from the Danish National Research Foundation (DNRF78) is gratefully acknowledged.
e-mail: akock@creates.au.dk. Address: Aarhus University and CREATES, Fuglesangs Alle 4, 8210 Aarhus V.