University student enrollment forecasts by analyzing structural ratios using ARIMA-methods

BY STEPHAN BOES AND PETER PFLAUMER*

SUMMARY: Forecasts for the number of students in Germany are conducted by the Kultusministerkonferenz. They use a transition model which does not allow for prediction intervals and therefore lack a measure of uncertainty of the forecast. Since the uncertainty is high for such forecasts, this lack is of importance. In this paper, structural ratios, relating the number of university students to the population of the same age, are analyzed and forecasted using ARIMA-models with outliers. Multiplying these ratios with official population forecasts for Germany provides the future number of students, additionally giving prediction intervals. This number will increase from 1.94 million in 2002 to 2.35 million in 2015. The uncertainty of the forecast is high; the forecast interval in 2015 will range between 1.72 and 2.98 million at a 95 % confidence level.

KEYWORDS: Higher education enrollment forecast, ARIMA-models, demographic development. JEL C22, I23, J11.

1. INTRODUCTION

The aim of our work is to give an estimate for the future number of students in higher education institutions in Germany, with a special focus on uncertainty, and thereby to give an alternative to the forecasts of the Standing Conference of the Ministers of Cultural Affairs of the Laender of the Federal Republic of Germany (Kultusministerkonferenz, in the following abbreviated as KMK). The KMK uses a transition model to forecast the number of students. It has been shown that the task is difficult, because the uncertainty is high (Boes and Pflaumer, 2002). But the KMK forecasts lack information on uncertainty.

We use data on student numbers and on population by age to calculate and forecast structural ratios. Making use of ARIMA methods, we are also able to show what the level of uncertainty is. The effects of reunification of Germany are to be modelled within the ARIMA framework. They are interpreted here as outlier effects.

In Section 2, we describe the data used in the analysis and motivate the assumptions necessary for the analysis. Section 3 outlines the ARIMA-analysis in the presence of outliers. The analysis and forecast of the data is shown in Section 4; Section 5 contains some concluding remarks.

Received: 26.07.04 / Revised: 26.07.05
* Supported by the German Research Foundation (DFG). We are grateful to an anonymous referee for some helpful comments.
2. DATA AND GENERAL ASSUMPTIONS

The data for student numbers are collected by the Federal Statistical Office (2002)\(^1\) since 1972 for Western Germany. Since 1992, these numbers give the sum of both Western and Eastern of the Federal Republic of Germany. They all relate to the start of the Winter Semester of the concerning year, i.e. to October 1st. The population data refer to December 31 of the concerning year, so there is a gap of three months between the two data sets. This difference is neglected because of low mortality within these age groups.

The data are given in groups of single age classes with the open classes 18 and younger and 39 and older at the low and high ends, but the main ages of studying within the German system are supposed to be between 19 and 25. Consequently, the German Federal Ministry of Education and Research gave a participation rate in Higher Education as the total number of students related to the population aged 19 to 25 until 2001. But since the older age groups (students older than 25) have grown in number mostly due to demographic reasons – especially in the 1990s, where the baby boom generation became older than 25 – the rate given appeared to be biased and was therefore omitted in the publication of 2003.

We take a different approach for student participation rate. We relate the number of students in the age group 19 to 25 (adding the younger students to this group, since their number is small) to the population aged 19 to 25, and build two groups of older students, aged 26 to 32 and 33 to 39, again related to the concerning age groups (with the whole open class 39 and older contained in the oldest group). These other age groups are set to the same length as the first, and cover the range given by the Federal Statistical Office. This so called structural ratio method has been supposed by Ahlburg (1985) and United Nations (1990), for example.

The analysis is done separately for traditional universities and Universities of Applied Sciences (UAS, in German: Fachhochschulen), because it must be assumed that the student behavior is different between these groups. Overall, there are six time series to be forecast. The analysis of these age groups should give an overview over the age distribution of students in Germany. It will also be shown what effect the demographic development has on the number of students, especially the ratio of older and younger ones.

One problem arising is the presence of an effect of reunification taking place in 1992, where the data became combined. Facing this, we try and interpret this effect as an outlier. We take an assumption by the KMK (1998, p. 4) that the student behavior will be about the same by the year 2000 in the old and new parts of reunified Germany. This assumption is supported by empirical evidence, since the proportion of school leavers qualified for university access who finally take up studies was already the same in 1999 in the overall view. As a further support to this assumption, the KMK

---

\(^1\) We want to thank the German Federal Statistical Office for kindly delivering the data.