World Patent Information 42 (2015) 19–27

Contents lists available at ScienceDirect

World Patent Information

journal homepage: www.elsevier.com/locate/worpatin

A dynamic log-linear regression model to forecast numbers of future filings at the European Patent Office

Peter Hingley [a, *], Walter Park [b]

[a] European Patent Office, Munich, Germany
[b] American University, Washington, DC, USA

ARTICLE INFO

Article history:
Received 30 January 2015
Received in revised form
1 July 2015
Accepted 10 July 2015
Available online 21 July 2015

Keywords:
Business cycles
Lognormal
Gross domestic product (GDP)
Patent filings forecasts
Research and development expenditure (R&D)
Linear model

ABSTRACT

An econometric model is applied to forecast future levels of patent filings at the European Patent Office out to 2019, using historical data from 1990 to 2013 with 28 source country terms. Descriptors include Research and Development expenditures and Gross domestic product, where the latter is split into trend and business cycles components. The model is applied to logarithmically standardised data.

The effects on the forecasts of additional future positive and negative stimuli to the GDP components are considered. Reasonable forecasting accuracy is found. Using a series of shorter historical data windows may give improved accuracy for short term forecasts.

1. Introduction

The European Patent Office (EPO) forecasts future patent filings in order to plan for likely workloads in the patent granting process, such as expected numbers of searches, substantive examinations, grants and renewals. These plans have implications for the requirements for staff and infrastructure. Procedurally, there is an annual cycle that proceeds from the forecasts via a business plan to finalisation in a budget document [1]. This budget is renewed annually and covers five years beyond the year in which it is produced.

The time series that are to be forecasted are shown in Fig. 1 with data up to 2013. EPO filings are a mixture of different types. Here we will consider forecasting the sum of Euro-direct filings and Euro-PCT international phase filings (Total filings in Fig. 1), after removing divisional filings (a form of retrospective Euro-direct filing that is forecasted separately). Other types of filings and downstream workload forecasts are then usually obtained by applying ratios to the forecasts for Total filings.

A variety of approaches are available that are based on historical data [1–3] or surveys [4]. The regression method that will be considered involves a dynamic log-linear model for annualised data that has been used since 2007 [5]. This operates on transformed EPO Total filings from 28 countries or regions, with autoregressive terms as well as source country Gross Domestic Product (GDP) and Research and Development expenditures (R&D) as independent variables. The model has recently been extended to consider the effects of business cycles [6]. This paper discusses the way that the approach has been customised for the forecasting process at EPO.

Matters of particular concern include transforming the data to achieve stationarity, how to calculate confidence intervals for the filings forecasts and how to interpret the forecasts and their accuracy against the later outcomes. The paper is organised as follows. Section 2 explains the model. In Section 3.1 a panel data set from 1990 to 2013 is fitted both to a model in levels and to a model in year-to-year differences. Section 3.2 shows the forecasts and interprets them. Section 4 considers the effect of a hypothetical boom or recession for one year during the forecasted period and also a scenario that is based on assumptions about the shape of the future business cycle. Section 5 looks at stability by fitting subsets of the same data in terms of a number of overlapping time windows. Section 6 discusses further directions.

2. The model for making parameter estimates and forecasts

The following regression model is used for EPO Total filings from a source country:-

* Corresponding author.
E-mail address: phingley@epo.org (P. Hingley).

http://dx.doi.org/10.1016/j.wpi.2015.07.002
0172-2190/

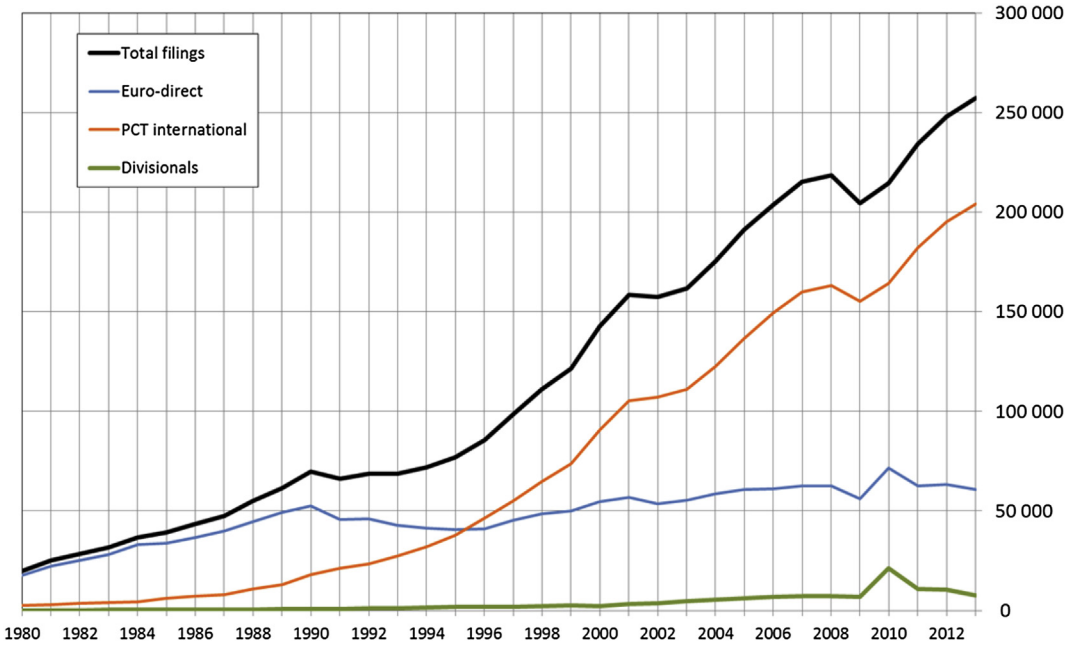

Fig. 1. The historical series of patent filings at EPO. (Adapted from Ref. [4]). Total filings are the sum of Euro-direct and PCT international phase from which divisional filings have been removed.

$$\log\left(\frac{P}{L}\right) = \alpha_0 + \alpha_1 \log\left(\frac{P}{L}\right)_{-1} + \alpha_2 \log\left(\frac{P}{L}\right)_{-2} + \alpha_3 \log\left(\frac{R}{L}\right) + \alpha_4 \log\left(\frac{Y^T}{L}\right) + \alpha_5 u + \varepsilon$$

Where P is the number of EPO Total filings from a source country[1];

L is the number of workers in the source country[2];
$_{-1}$ and $_{-2}$ indicate lags of one year and two years respectively;
R is R&D expenditures,[3] usually lagged by 5 years;
The GDP of the source country Y[4] is split into two components:-
Y^T is the "trend" level of output
u is the business cycle variable (a ratio of cyclical GDP to trend GDP);
ε is an error term, assumed to be normal with constant variance;
log( ) denotes natural logarithm.

Total filings P are transformed as indicated to log(P/L). This allows for a standardisation between countries, as L is treated as a proxy for country size, and for stabilising error by the logarithmic transformation. Based on [10], the value of R is lagged by five years in order to incorporate the concept that R&D expenditures have their effect after a delay. Most EPO filings are subsequent filings that take place up to a year after first filings, so the assumption is that R&D expenditures "cause" first filings about 4 years later on. Qualitatively similar results are obtained via a model with no lag in R [6], and in Section 5 below some comparisons are made between lags of 1, 3, 5 and 7 years. No time dummies are included, which gives better forecasting ability by assuming that the process remains stable over time.

The GDP term Y is decomposed via the Hodrick and Prescott filtering method [11] into its trend and cyclical components (Y^T and Y^C respectively) and then the business cycle variable is $u = Y^C/Y^T$. This is detailed in Ref. [6], where it is demonstrated that the usage of u and Y^T rather than Y improves the goodness of fit to the model for filings on the historical training data set, and so may also provide improved forecasting ability.

Annex 1 indicates the way that the forecasts for EPO filings from the source countries and their variabilities were calculated and combined to make the forecasts for Total filings. The authors will be prepared to share further details of the methods on request.

3. Results

3.1. Fitting the models

The analysis here reflects the data up to 2013 that were available in the second half of 2014. The model is fitted to a 28 source country-of-origin data set using annualised EPO Total filings from 1990 to 2013.[5] Data for the variables are calculated both as levels

[1] Filings refer to the sum of Euro-direct and PCT international phase filings [1], excluding divisionals, except where otherwise specified. Euro-direct are obtained from the EPO production database and PCT are as reported by WIPO.

[2] Number of workers data are provided by the World Bank [7].

[3] R&D expenditures are business enterprise research and development expenditures (BERD) from OECD MSTI 2013 edition 2 [8], at constant 2005 PPP international dollars. Comparable data are taken from UNESCO for countries that are not given by MSTI. For most countries, data were available up to 2012 at the time of analysis and have been trended out to 2013 and beyond by using linear regression on the last 10 years of available data.

[4] GDP expenditures are obtained from the World Bank's World Development Indicators [7], or Penn World Tables [9] for Chinese Taipei, standardised to real constant 2005 PPP international dollars. Agency forecasts for 2014 and 2015 are used where available and for later years have been trended by using linear regression on the last 10 years of available data.

[5] 27 individual countries were the following, together with a 28th group "ZZ" that represented the residual between the measured Total filings in a year and the sum from the 27 countries: Australia (AU), Austria (AT), Belgium (BE), Brazil (BR), Canada (CA), China & Hong Kong (CN-HK), Denmark (DK), Finland (FI), France (FR), Germany (DE), Hellas (GR), Ireland (IE), Israel (IL), Italy (IT), Japan (JP), Republic of Korea (KR), The Netherlands (NL), New Zealand (NZ), Norway (NO), Portugal (PT), Singapore (SG), Spain (ES), Sweden (SE), Switzerland (CH), Chinese Taipei (TW), United Kingdom (GB), United States of America (US), Others (ZZ).