Forecasting Patent Filings at the European Patent Office (EPO) with a Dynamic Log Linear Regression Model: Applications and Extensions

Peter Hingley and Walter Park

Abstract We describe the application of a structural econometric model to forecast annual and quarterly total filings of patents at the EPO, amid cyclical shocks. Such models are applicable for national, regional, and international patent offices for budgetary planning purposes. We also provide out-of-sample forecasts out to 2019. The exercise predicts strong further growth in EPO filings from China. The forecasts appear largely optimistic out to 2019. Other findings are that quarterly data do not exhibit as much sensitivity to cyclical influences as do annual data, consistent with the view that innovation decisions are made based upon longer term factors. Furthermore, filings respond more directly to market fluctuations rather than indirectly to deviations in research and development (R&D) spending from trend. The estimates also vary by technological field. Further possibilities for an extension of the approach are discussed, including the usage of country-to-country correlations of forecasts to unearth communalities in patenting behaviour between countries.

Keywords Total filings (Euro-direct filings plus Euro-PCT international phase filings) · Lognormality · Business cycles · Gross domestic product (GDP) · Patent offices · Research and development expenditure (R&D) · Linear model · Variance

1 Introduction

Planners at the European Patent Office (EPO) make forecasts for future patent filings in order to power business planning for all the major points of the patent granting process, such as searches, substantive examinations, grants and renewals.

P. Hingley
European Patent Office, Munich, Germany
e-mail: phingley@epo.org

W. Park (✉)
American University, Washington D.C., USA
e-mail: wgp@american.edu

There is an annual cycle that proceeds from the forecasts via the business plan that is finalised in a budget document (Hingley and Nicolas 2006). Academic scholars are interested in patent forecasts in order to better understand the determinants of innovation and productivity, which are fundamental influences on economic growth and human progress.

The purpose of this study is to model EPO filings behaviour and derive the implications for forecasting EPO filings out-of-sample, over the medium and longer term. The motivation for this is a study on forecasting EPO filings by Park (2006), in which only long term trends were taken into account. The report was conducted amid an economic boom around the world, driven by the housing market, relatively low energy prices, and the expansion of multinational firm activity, such as outsourcing and offshoring, particularly by the leading technology companies. The earlier study did not anticipate, nor allow for, the effects of a major economic downturn that has come to be known as the *Great Recession*. Consequently Park (2006) produced rather optimistic projections of real gross domestic product (GDP) and EPO filing activity. Instead, there were decreases in EPO filings by leading source countries, like the U.S., Japan, and Germany, particularly in 2009, the critical year of the Great Recession. Thus, business fluctuations seem to have some effect on patenting. However, there also was a relatively quick recovery in filings during 2010–2011. Moreover, for the other countries, their patent filings changed relatively little during the period of the global financial crisis. China's filings actually rose during the Great Recession period. Thus, this raises the issue of how significantly business fluctuations affect patenting and matter to forecasts of future patenting. What is the direction of effect: is patenting pro-cyclical (moves positively with business cycles) or counter-cyclical (moves inversely with business cycles)? How much do business cycles really contribute to changes in patenting? How lasting are the effects of business cycle shocks on patenting? What is the value-added of forecasting business cycles to forecasting EPO filings and applications?

A variety of econometric regression based approaches exist.[1] One of those methods will be considered here, that is based on a dynamic log-linear model (see Hingley and Park 2015a; Park 2006). In an endeavour to improve the model, a business cycles component has been included to supplement trend GDP.

Business cycles refer to fluctuations of output around some average or long run trend path. The trend path of output is usually determined by productivity, technology, or resources. The deviations of output from trend are usually due to nominal shocks, such as aggregate demand shifts in the presence of wage and price rigidities. The reason business cycles are a concern is that when output is below trend, resources are underutilized and unemployment may increase. When output is above trend, inflationary pressures tend to arise. However, there is much debate

---

[1]See Hingley and Nicolas (2004, 2006), Hidalgo and Gabaly (2012), and Meade (2006) for historical data analyses and European Patent Office (2013) for the survey approach.

about how long lasting business cycles are. One school of thought (for example, the old-Chicago school) is that markets self-correct relatively quickly and that most of the output fluctuations are equilibrium movements or shifts in trend output itself. Another school (for example, the old-Keynesian school) is that economies experience persistent disequilibria. The findings in this chapter present a view that is intermediate between these two schools: output deviations from trend are neither permanent nor equilibrium phenomena.

The chapter is organized as follows. The next section outlines our empirical framework. Section 3 presents estimates of the model and some out-of-sample forecasts. Section 4 re-examines the model from some alternative perspectives, and Sect. 5 discusses other applications. Section 6 concludes. Overall, the study demonstrates the desirability and feasibility of forecasting EPO patent filings under uncertainty about future movements in GDP or the market outlook. The significance is that, to the extent that innovation activities are attuned to market conditions, market fluctuations should affect the capacities and incentives to seek patent protection. Furthermore, the frequency of cycles matters. EPO filings are affected by business cycles when yearly data are considered. For shorter data frequencies—namely quarter to quarter—EPO filings are not as sensitive to business fluctuations. A reason discussed is that significant innovation decisions are not likely to be made on the basis of short-lived circumstances. But on a yearly scale, fluctuations can affect budgetary and other cost factors that impinge upon innovation plans.

2 Empirical Framework

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

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

where P is the number of EPO filings filed by a source country,[2] L is the number of workers in the source country,[3] subscripts -1 and -2 indicate lags of one year and two years respectively, R is R&D expenditures,[4] usually lagged by 5 years, and ε is

---

[2]EPO filings refers to the sum of Euro-direct and Patent Cooperation Treaty (PCT) international phase filings (see Hingley and Nicolas 2006), excluding divisional filings. Euro-direct are obtained from the EPO production database and PCT are as reported by the World Intellectual Property Organization (WIPO).

[3]Number of workers data are provided by the World Bank *World Development Indicators*: http:// data.worldbank.org/data-catalog/world-development-indicators.

[4]R&D expenditures are business enterprise research and development expenditures (BERD) from OECD MSTI 2013 edition 2 (see OECD 2014), at constant 2005 PPP international dollars. Comparable data are taken from UNESCO for countries that are not given by MSTI.

an error term, assumed to be normal with constant variance. ln denotes natural logarithm. The GDP of the source country (Y) is split into two components: Y^T the "trend" level of output, and u a business cycle indicator (namely, the ratio of cyclical GDP to trend GDP).[5] The Hodrick and Prescott (2007) filter method was used to separate cycles from trend (and is further explained in Hingley and Park 2015b).

Filings P are transformed as indicated to $\ln(P/L)$. This allows for a standardisation between countries because L is a proxy for country size, and for stabilising the error by using the logarithmic transformation. Based on Ditka (2006), the value of R is lagged by five years in order to incorporate the concept that R&D expenditures take time to produce patentable inventions. The effects of possible booms and/or recessions within the forecast period can be assessed by manipulating u. Table 1 shows some sample statistics associated with the calculated u. The table illustrates that business cycles vary by country. Of course, there are common global shocks. But not all of these shocks are transmitted to national economies in the same way. Some economies are better insulated against external shocks, depending upon on their policy or institutional regimes. Furthermore, different countries may be in different phases of the business cycle. Hence, our decomposition method has been applied separately to each of the countries in the sample.

The model is fitted over a group of countries that includes a rest-of-the-world (ROW) class. We forecast total filings usually up to 6 years beyond the training data period. In order to obtain the forecasts for filings in a future year, the individual forecasts derived from the model by country are summed across countries. At the level of filings P, the assumed distribution of the error ε, and hence also that of P itself, is assumed to be lognormal. The technique for estimation takes this into account. Other reported studies have often not done this kind of transformation and have therefore worked with imprecise confidence intervals. Let $v = P/L$, with log (v) distributed $N(\mu, \sigma^2)$, a normal distribution with mean μ and variance σ^2. The model contains fixed effects—a separate intercept for each of 28 countries (α_{01}, α_{02}, ..., α_{028})—and common slope parameters ($\alpha_1, \alpha_2, \alpha_3, \alpha_4$ and α_5). The linear model is fitted to the transformed data for the various countries simultaneously to determine the estimates of $\hat{\mu}$ and $\hat{\sigma}^2$ for each country. This is done in the usual way by gathering the data underlying the independent variables, including the {0, 1} dummy variables for the intercepts, into an ($n \times p$) design matrix Z. The parameters to be estimated are themselves stacked into a ($p \times 1$) parameter vector B. Let ^{T} indicate transposition:

---

[5]GDP expenditures are obtained from the World Bank's *World Development Indicators* and *Penn World Tables* https://pwt.sas.upenn.edu for Taiwan, standardised to real constant 2005 PPP international dollars.

Table 1 Sample statistics of the business cycle variable, u

Country	Mean of 'u'	Standard deviation	Coefficient of variation	Minimum	Maximum
Australia	−0.0041	1.0368	−252.0	−3.006	2.389
Austria	−0.0038	1.0803	−281.5	−2.357	2.319
Belgium	−0.0042	0.9825	−235.8	−2.076	2.229
Brazil	−0.0186	2.3141	−124.7	−5.733	6.845
Canada	−0.0091	1.3563	−149.2	−2.625	2.551
China incl. Hong Kong	0.0239	1.7829	74.6	−4.102	4.748
Denmark	−0.0087	1.3356	−153.6	−3.213	2.823
Finland	−0.0274	2.1292	−77.7	−4.936	4.864
France	−0.0049	0.9082	−187.1	−2.161	1.828
Germany	−0.0051	1.2824	−249.5	−3.634	2.369
Greece	−0.0231	1.4292	−61.9	−3.931	2.798
Ireland	−0.0239	1.7995	−75.3	−3.542	5.188
Israel	−0.0088	1.4782	−168.5	−2.747	4.436
Italy	−0.0050	1.1286	−223.8	−3.133	2.315
Japan	−0.0073	1.2073	−165.8	−3.868	2.329
Korea, Republic	−0.0025	2.0373	−802.1	−6.688	4.039
Netherlands	−0.0063	1.1360	−179.3	−1.772	2.619
New Zealand	−0.0117	1.1756	−100.4	−3.343	1.880
Norway	−0.0071	1.1159	−157.8	−2.527	2.328
Portugal	−0.0217	1.6022	−73.9	−2.746	2.933
Singapore	−0.0110	2.4053	−217.8	−5.976	5.154
Spain	−0.0140	1.0673	−76.0	−1.667	2.389
Sweden	−0.0102	1.5008	−146.9	−5.095	2.964
Switzerland	−0.0065	1.0879	−167.2	−1.695	2.693
Taiwan	0.0016	1.6706	1068.9	−5.143	3.414
United Kingdom	−0.0133	1.2229	−92.0	−2.612	2.756
United States	−0.0084	1.2254	−145.2	−3.400	1.860
World_ROW	−0.0059	1.1772	−199.4	−2.738	2.159

Notes $u = 100 \times$ (cycle/trend), where GDP = trend + cycle; i.e., using the Hodrick-Prescott filter, real GDP has been decomposed into its trend and cylical components

$$B^T = [\alpha_{01}, \alpha_{02}, \ldots, \alpha_{028}, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5]$$

$$
Z = \begin{bmatrix}
1 & 0 & \ldots & 0 & a_{1,1} & a_{2,1} & a_{3,1} & a_{4,1} & a_{5,1} \\
1 & 0 & \ldots & 0 & a_{1,2} & a_{2,2} & a_{3,2} & a_{4,2} & a_{5,2} \\
. & . & . & . & . & . & . & . & . \\
1 & 0 & \ldots & 0 & a_{1,24} & a_{2,24} & a_{3,24} & a_{4,24} & a_{5,24} \\
0 & 1 & \ldots & 0 & a_{1,25} & a_{2,25} & a_{3,25} & a_{4,25} & a_{5,25} \\
0 & 1 & \ldots & 0 & a_{1,26} & a_{2,26} & a_{3,26} & a_{4,26} & a_{5,26} \\
. & . & . & . & . & . & . & . & . \\
0 & 1 & \ldots & 0 & a_{1,48} & a_{2,48} & a_{3,48} & a_{4,48} & a_{5,48} \\
. & . & . & . & . & . & . & . & . \\
. & . & . & . & . & . & . & . & . \\
. & . & . & . & . & . & . & . & . \\
0 & 0 & \ldots & 1 & a_{1,792} & a_{2,792} & a_{3,792} & a_{4,792} & a_{5,672}
\end{bmatrix}
$$

where $a_{i,j}$ is the value of the independent variable i for the jth observation in the dataset. The parameter estimates $\hat{B}$ are calculated by least squares and the associated error variance of ε is calculated as $\hat{\sigma}^2$:

$$\hat{B} = Z \cdot (Z^T Z)^{-1} \cdot Z^T \cdot \log(v)$$

$$\hat{\sigma}^2 = (\log(v) - Z\hat{B})^T \cdot (\log(v) - Z\hat{B})/(n - p)$$

where $\log(v)$ is the $(n \times 1)$ vector $\log(v_{1,1}, v_{1,2}, \ldots, v_{1,24}, v_{2,1}, \ldots, v_{28,672})^T$, that is the string of transformed observations (or differences in transformed observations) from each country by years in the training set, laid out with years within countries repeating faster than countries.

On the logarithmic scale, the fitted values of the observations are given by the matrix inner product $Z\hat{B}$ within the training set. The forecasts for each country for each future time point are given by projecting further $(1 \times p)$ rows z that are equivalent to rows of Z but taking independent variables beyond the data set and calculating the inner product $z\hat{B}$.

The lognormal distribution of v has mean $\gamma = \exp(\mu + \sigma^2/2)$ and variance $\text{Var}[\gamma] = \gamma^2 (\exp(\sigma^2) - 1)$ (see Johnson et al. 1994). The fitted values on the scale of v are therefore taken from the linear model as $\hat{\gamma} = \exp((z\hat{B}) + \hat{\sigma}^2/2)$. The estimated number of filings from a country at a given time point is then $w = L\hat{\gamma}$, with an estimated variance, $\text{Var}[w] = L^2 \hat{\gamma}^2(\exp(\hat{\sigma}^2) - 1)$.

It should be noted that, during the forecast period beyond the training data set, it is necessary to forecast L and also the independent variables Y^T, u and R.[6] This is usually done by straight line regression projection from the 10 most recent available years of data in the training set, and is a source of additional variability for the forecasts. The goal is to forecast Total filings as the sum of the forecasted filings per country of origin. This is $\hat{w}_{TOTAL} = \sum w = \sum L_i \hat{\gamma}_i$, where $\sum$ indicates summation over countries. $\hat{\gamma}_i$ is the filings estimate for country i.

In the following, Ξ is the (28×28) covariance matrix between countries on the log scale, that is estimated from the linear model by $\hat{\Xi} = (Z^T Z)^{-1} \hat{\sigma}^2$. Two alternative approaches are taken to estimate the variance of $\hat{w}_{TOTAL}$:

1. The sum of the estimated variances for the countries.

$$\text{Var}[\hat{w}_{TOTAL}] = \text{Var}\left[\sum_i L_i \cdot \hat{\gamma}_i\right] = \sum_i L_i^2 \cdot \hat{\gamma}_i^2 \left(\exp(\hat{\Xi}_{ii}) - 1\right)$$

The summation is over all countries i: 1, …, 28.

2. The sum of the estimated covariances for all pairs of countries i and j.

$$\text{Var}[\hat{w}_{TOTAL}] = \text{Var}\left[\sum_i L_i \cdot \hat{\gamma}_i\right] = \sum_i L_i \sum_j L_j \cdot \hat{\gamma}_i \cdot \hat{\gamma}_j \left(\exp(\hat{\Xi}_{ij}) - 1\right)$$

This uses a formula for the variance of a sum that is based on an extension of the formula for the variance of the mean of the lognormal distribution (Soderlind 2013). The summation is over all country pairs i and j. The country pairs include each country paired with itself, in which case the covariance term is the same as the term that feeds into expression 1 above.

Approach 1 is slightly easier to calculate than approach 2, but the covariances are likely to be relevant effects, since the model is fitted simultaneously over countries with some common parameters between countries, with international common developments affecting patent filings in many countries at the same time, and because of possible non-stationarity of the historical training set data (Ditka 2006). Approach 2 helps to take account of these factors. Here $\text{Var}[\hat{w}_{TOTAL}]$ will be calculated by both approaches. The estimates will be compared to get an idea of stability of the model and to assess the attempts to assure stationarity by differencing both the variables.

As was already mentioned, a linear trend approach is used to forecast the independent variables Y^T, u, and R, that are then included as inputs to Z in order to use the estimated parameter matrix $\hat{B}$ to find the fitted values $\hat{\gamma}_i$ for each country i for each forecasted year outside the training data set. In the case of R, the assumed

---

[6]Forecasts for Y^T and u are obtained from the HP filter. Since R is lagged 5 years, observed historical values can be used for the first few future forecasted years before trending is required.

5 year lag means that historical known values can be used for the forecasts of the next few years in the future. Since the model includes autoregressive terms that relate to lags of filings at one and two years, the forecasts for more than two years out themselves use inputs of filings forecasts from one and two years previously. These inputted forecasts are subject to variability by the same error process, and the other inputs are also subject to error because of the usage of a trend to estimate them. Therefore it is likely that $\mathrm{Var}[\hat{w}_{TOTAL}]$, as given directly either by method 1 or 2 above, is not great enough to cover all the variability that is inherent in this approach. In order to cope with this to some extent, a pragmatic compound variance method is used. The variance of the filings forecast for a future year is taken as the sum of the variances, by either method 1 or 2, taken over all the forecasted years up to and including s:

$$\mathrm{Var}[(\hat{w}_{TOTAL})_s] = \sum_{l=1}^{s} \mathrm{var}(\hat{w}_{TOTAL})_l$$

From this variance, 95 % confidence limits for the forecast of total filings in a future year are calculated by the usual normal assumption, $(\hat{w}_{TOTAL})_s \pm 1.96*$ $\mathrm{SE}[(\hat{w}_{TOTAL})_s]$, where SE indicates standard error and is the square root of $\mathrm{Var}[(\hat{w}_{TOTAL})]$. The confidence limits are appropriate for the predicted values of the mean. The mean is forecasted because the process uses essentially unchanging historical training data for all years up to the last year in the data set, with only one added data point per country in each successive annual forecasting exercise.[7]

3   Results

3.1   Model Estimates

The model is fitted to a 28 source country-of-origin data set using their annualised EPO total filings from 1990 to 2013.[8] The linear model is fitted both to the levels

---

[7]Alternative confidence interval/prediction interval formulae are given in the literature (Draper and Smith 1981), depending whether the intervals are sought for the mean or for an individual new observation.

[8]27 individual countries were the following, together with a residual 28th group "ZZ" that represented the residual between the measured total filings in a year and the sum from the 27 countries, as described at European Patent Office (2015): 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), Taiwan (TW), United Kingdom (GB), United States of America (US), Others (ZZ). The analysis reported here was done in mid 2014, before GDP data were updated to give results in Hingley and Park (2015a).

Table 2 Estimates of the log-linear model

Parameter		Model in levels		Model in differences	
		Estimate	Standard error	Estimate	Standard error
α_1	AR1	0.823	0.038	−0.072	0.039
α_2	AR2	0.043	0.036	−0.045	0.037
α_3	R	−0.013	0.027	−0.050	0.103
α_4	Y^T	0.444	0.075	2.145	0.395
α_5	u	0.860	0.414	1.109	0.356
Error variance		0.0228		0.0247	
Data point standard deviation		0.151		0.157	

Notes Model for EPO filings with training data set from for Levels (1990–2013) and Differences (1991–1990 to 2013–2012). Country fixed effects are included (not shown). AR1 and AR2 are one year and two year lags respectively for standardised patent filings, R is standardised R&D expenditure by business sector, Y^T is the trend level of standardised GDP, u is the business cycle variable). Also shown are Error variance and its square root, data point standard deviation. See Footnote 8 for the key to the country names

data and to the first-differences data. The parameter estimates and standard errors obtained by both approaches are shown in Table 2. Following Park (2006), the intercept α_0 is allowed to vary from country to country while the other parameters α_1 to α_5 are considered as common over countries so that pooled estimates are obtained The 5 slopes appertaining to α_1 to α_5 are shown in Table 2. The standard deviations of a data point estimate at the bottom of the table are similar for the two approaches. In terms of formal significance testing, several of the parameter estimates are not statistically different from zero under an assumption of normality, in that the absolute value of the estimate is more than 1.965 times its estimated standard error.

For the model in levels, all parameter estimates are significant by this approximate test except for the second order autoregression (AR2) and R. But for the model in differences, all the country intercepts are not significant except those for BR, CN-HK, KR, NZ, PT, SG and ES. AR1, AR2 and R are technically not significant as well, although AR1 is almost significant. However, we believe that all the included variables are useful structural descriptors of the process and should remain in the model to assist in making the forecasts. The structural dependency of the model on the parameters is manifested for the model in levels, where the non-stationary process is adequately described. What is not guaranteed for levels is the extent to which the set of independent variables is causal for the filings process.

For the model in differences, the significance is also assessed at the level of differences, and it may be that effects are in fact significant when the data are transformed back to levels. We do not succumb to the temptation to remove parameters in order to obtain a reduced model where every remaining parameter is statistically significant. There are 672 observations for the model in levels and 644

observations for the model in differences, so the removal of 33 degrees of freedom
for estimation does not seriously degrade the quality of the residual sum of squares
that is used to estimate the error terms.

At least for the model in differences, no particular importance should be given to
the sign (positive or negative) of a parameter estimate, even when it is significant.
A process by which differences of an independent variable have an effect on dif-
ferences in filings, particularly after transformation, can look negative when it
would in fact be positive back on the scale of levels.

Common economic factors that affect patenting in several countries at the same
time, as well as the pooling of parameter estimation over the countries, induces
some correlations between the parameter estimates. Table 3 shows simplified
parameter correlation matrices for both levels and differences. Since the 28 inter-
cepts behaved rather similarly to each other in terms of correlations, examples are
shown only for countries from the four important sources: China, Germany, Japan
and US.

Table 3 Correlation coefficients for parameter estimates

Models in levels								
Parameter	CN-HK	DE	JP	US	AR1	AR2	R	Y^T	u
CN-HK	1	0.996	0.996	0.996	0.260	0.068	0.100	−0.951	0.072
DE	0.996	1	0.999	0.999	0.254	0.066	0.085	−0.953	0.072
JP	0.996	0.999	1	0.999	0.254	0.069	0.074	−0.949	0.071
US	0.996	0.999	0.999	1	0.255	0.068	0.075	−0.949	0.072
AR1	0.260	0.254	0.254	0.255	1	−0.910	0.004	−0.224	−0.041
AR2	0.068	0.066	0.069	0.068	−0.910	1	−0.188	−0.008	0.046
R	0.100	0.085	0.074	0.075	0.004	−0.188	1	−0.376	0.067
Y^T	−0.951	−0.953	−0.949	−0.949	−0.224	−0.008	−0.376	1	0.067
u	0.072	0.072	0.071	0.072	−0.041	0.046	0.067	0.067	1
Models in differences								
Parameter	CN-HK	DE	JP	US	AR1	AR2	R	Y^T	u
CN-HK	1	0.093	0.093	0.130	−0.138	−0.137	−0.283	−0.615	0.002
DE	0.093	1	0.020	0.022	−0.035	−0.038	−0.057	−0.093	0.006
JP	0.093	0.020	1	0.025	−0.072	−0.070	−0.103	−0.075	0.001
US	0.130	0.022	0.025	1	−0.036	−0.039	−0.052	−0.152	0.005
AR1	−0.138	−0.035	−0.072	−0.036	1	0.060	0.028	−0.111	0.017
AR2	−0.137	−0.038	−0.070	−0.039	0.060	1	−0.003	−0.092	0.029
R	−0.283	−0.057	−0.103	−0.052	0.028	−0.003	1	−0.027	0.034
Y^T	−0.615	−0.093	−0.075	−0.152	−0.111	−0.092	−0.027	1	−0.047
u	0.002	0.006	0.001	0.005	0.017	0.029	0.034	−0.047	1

Notes From the model for EPO filings with the training data set in Levels and First-Differences. Country
intercept correlations are shown for China, Germany, Japan and US, and are similar for the other 24
countries (See Table 2 for keys to parameter symbols)

Table 4 Total filings forecasts by the model in levels and first-differences

Year	Actual total filings	Model in levels			Model in differences		
		Total filings forecast	Standard error 1	Standard error 2	Total filings forecast	Standard error 1	Standard error 2
2013	257,457	261,929	(3105)	(3487)	263 342	(3144)	(3204)
2014		271,987	3256	3699	267,554	4520	4832
2015		287,637	4806	5510	286,476	5586	5935
2016		304,826	6184	7181	307,337	6603	7010
2017		323,868	7561	8874	331,276	6751	8094
2018		344,939	8036	10,694	358,284	7914	9232
2019		368,272	10,683	12,715	388,887	9127	10,461

Notes Standard error 1 and Standard error 2 refer to the square roots of variance methods 1 and 2 respectively for $\text{Var}[\hat{w}_{\text{TOTAL}}]$, as explained in Sect. 2. The results for 2013 show the model fit to the last year in the training data set, with standard errors in brackets. The cumulation of variances starts in 2015

For the model in differences, 21 out of the 36 distinct correlation coefficients have negative signs, while for the model in levels only 9 have negative signs. In general, the correlation coefficients between parameters are far less significant for the model in differences than for the model in levels. This is an indicator of stationarity for the model in differences. Looking more deeply at the model in differences, the correlations between country intercepts are positive but generally not significant, although there are relatively high correlations between CN-HK and US; and between CN-HK and KR (0.340 not shown), although not especially between KR and US (0.083 not shown). R has negative correlations with country intercepts, including quite a high negative value with CN-HK and with KR (−0.175 not shown). The EPO filings from CN-HK and KR have been growing more strongly over the period than from most other countries, which may have led to these effects. Rapid development of EPO patenting from CN-HK and KR may also be reasons for their significant intercepts in Table 2 in terms of the model in differences.

3.2 Model Forecasts

Table 4 shows the fitted values or forecasts for total filings $\hat{w}_{\text{TOTAL}}$, for years 2014–2019, together with estimates of their variances $\text{Var}[\hat{w}_{\text{TOTAL}}]$ by the two methods. The forecasts depend on the future levels of the independent variables R, Y^T and u that are assumed.[9]

---

[9]For more technical details about the methodology, see Hingley and Park (2015a).
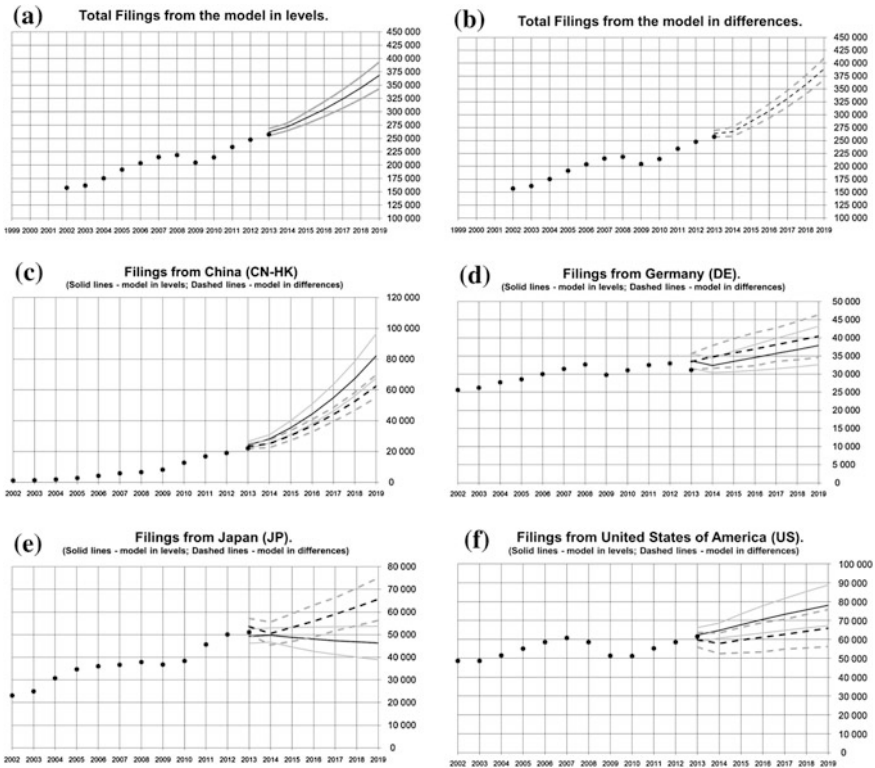
Fig. 1 Filings forecasts by the model in levels and the model in differences. *Notes Black lines* are the forecasts and grey lines are the 95 % confidence intervals for the forecasts. **a** Total filings by model in levels. **b** Total filings by model in differences. **c** Filings from China. **d** Filings from Germany. **e** Filings from Japan. **f** Filings from United States of America

Figure 1 shows a comparison between the actual and predicted values in levels and differences for the years 2014–2019. The total filings numbers are reported first, and then the filings from the important countries of origin CN-HK, DE, JP and US (as in Table 3). 95 % confidence intervals are calculated for each individual forecasted year, using the second variance method discussed in Sect. 2, and the forecasts and limits are connected over time by smoothed lines using Excel. Both models give optimistic forecasts for total filings. The model in levels proposes 368,272 in 2019, a compound annual growth rate of 6.1 % from 257,747 in 2013. The model in differences proposes 388,887 in 2019, which represents a compound annual growth rate of 7.1 % from 2013. Both models suggest low growth from 2013 to 2014 however (1.7 % for the model in levels and 3.9 % for the model in differences). Thereafter the two models roughly agree until about 2016, after which point the model in differences starts to move ahead. Forecasts towards the end of the period are of course very uncertain, so details of differences between the models out there should not be over-interpreted.

The confidence intervals widen towards the end of the forecasting period in both models. Standard error 1 is always less than standard error 2, which may indicate a tendency for positive covariances between forecasted filings for countries. At the beginning of the period, the relative sizes of the standard errors 1 and 2 are more similar for the model in differences than for the model in levels (in 2014 the ratio is 1.14 for levels and 1.07 for differences), but this situation is reversed at the end of the period (in 2019 the ratio is 1.19 for levels and 1.15 for differences). This may be related to the improved stationarity of the times series when using the model in differences.

Regarding the country forecasts in Fig. 1, there are some variations between the models for levels and differences, with levels giving higher forecasts than differences for China and United States, but differences giving higher forecasts than levels for the other two countries that are shown. For Japan (Fig. 1e), the forecasts for the model in levels decline while those in differences increase after experiencing a downward "kink" in 2014. However, this difference may not be significant because of the overlap between the confidence intervals until 2019. The strong level of projected growth for China in Fig. 1c should be noted, where about 39,000 additional filings per year are expected by 2019 compared to 2013, according to the model for differences, which makes a strong contribution to the overall increases of total filings that are envisioned in Fig. 1a, b.

4 Alternative Perspectives

This section briefly addresses some open issues, with the goal of checking for robustness and of suggesting extensions to this study. In particular, it examines three alternative ways to study the effects of business fluctuations on patenting filings at the EPO. The first is to examine quarterly data. The second is to capture cyclical movements in R&D. The third is to examine data by technological field (or sector).

Table 5 reports the results from estimating the EPO filings per-worker model on quarterly data. Column 1 shows the results for the period 1978–2011 and column 2 for a truncated period. For the quarterly data analysis, seasonal effects are controlled for by using quarterly dummies. Quarterly data are relatively noisier than annual data. Cyclical fluctuations may be small year to year, but within each year, there may be quite a bit of economic disturbances from month to month or quarter to quarter. However, what the results seem to indicate is that while patenting can vary with the business cycle, it does not vary strongly with short frequency business cycles. The 'u' variable is insignificant at explaining EPO filings.

The reason that short frequency business cycles are not a strong determinant of EPO filings is that quarter-to-quarter disturbances are relatively transient and unpredictable. It does not seem prudent for businesses to make innovation decisions based on such events. The objective of innovation is to develop new products and processes to improve efficiency or productivity, maintain or expand market share,

Table 5 Main regression model estimated using quarterly data

Sample period	(1)	(2)
	EPO filings	
	1978–2011	1978–2011
L.pl	0.400***	0.354***
	(0.023)	(0.025)
L2.pl	0.339***	0.339***
	(0.022)	(0.025)
yl	0.435***	0.255**
	(0.110)	(0.116)
rl	0.108***	0.195***
	(0.026)	(0.032)
u	0.149	0.140
	(0.099)	(0.104)
year	0.004**	0.005**
	(0.002)	(0.002)
1st Quarter		−9.480**
		(4.486)
2nd Quarter	0.093***	−9.392**
	(0.014)	(4.486)
3rd Quarter	−0.004	−9.487**
	(0.013)	(4.486)
4th Quarter	0.117***	-9.361**
	(0.014)	(4.485)
Country fixed effects	Included	Included
Observations	1590	1431

Dependent Variable: pl = ln(patents/worker)

Standard errors in parentheses: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$

Notes pl denotes natural log of patents per worker, L one-period lag, $L2$ two-period lag, yl denotes real GDP per worker, rl real business enterprise R&D (BERD) per worker, 'year' denotes the time trend, and 'u' the business cycle measure where u = cycle/trend of GDP. 1st, …, 4th quarter are dummy variables for their respective quarters

and strengthen the brand and core competencies of the firm. Innovation therefore typically targets a firm's long run profitability, which is why patenting and innovation are expected to depend, in general, on longer run factors, but as was seen with the results earlier with the annual data and discussion of the theory, year-to-year business cycles can nevertheless affect the availability of resources for innovation (or potentially the opportunity cost of innovation)—hence the sensitivity of annual patent filings to long frequency business cycles.

But at the quarterly data level, the time interval is fairly short. Firms typically seek other means to gain or maximize shorter run profits, such as financial or portfolio

investments, derivatives, swaps, pricing strategies (such as discounts, non-linear pricing, price discrimination), or employment strategies (using contractors and temporary workers). Firms generally do not use R&D investments to manipulate and influence short run profits. Principally, R&D and patenting strategies are for time horizons longer than a few quarters—hence the reason why business cycles are more likely to affect patenting across years rather than across quarters.

The next issue is whether other aspects of the model also undergo cyclical shocks. In the analysis thus far, only GDP was separated into its trend and cyclical components. It may also be possible for business R&D expenditures to have a cyclical and a trend component. This was investigated, and the results are shown in Table 6. The table introduces two new variables upon applying the Hodrick-Prescott (HP) filter to R&D: trend R&D per worker and the ratio of cyclical R&D to trend R&D (uR). The uR variable is statistically insignificant at conventional levels, whether the sample period is 1978–2011 or the shorter one of 1990–2011. It is also insignificant whether we omit or control for business cycles in GDP. Thus, for now,

Table 6 Extensions: cyclicality of research and development

Sample period	(1)	(2)	(3)	(4)
	1978–2011	1990–2011	1978–2011	1990–2011
L.pl	0.685***	0.783***	0.684***	0.783***
	(0.030)	(0.040)	(0.030)	(0.040)
*L*2.pl	0.131***	0.043	0.132***	0.044
	(0.025)	(0.038)	(0.025)	(0.038)
ln (GDP per worker)	0.456***	0.475***		
	(0.078)	(0.089)		
ln (Trend GDP per worker)			0.444***	0.449***
			(0.079)	(0.093)
u			0.768	0.928**
			(0.597)	(0.458)
ln (Trend R&D per worker)	0.078**	0.066*	0.080**	0.071*
	(0.037)	(0.037)	(0.037)	(0.037)
uR	−0.005	0.216	−0.051	0.158
	(0.202)	(0.146)	(0.219)	(0.157)
year	0.001	0.001	0.001	0.001
	(0.002)	(0.002)	(0.002)	(0.002)
Country fixed effects	Included	Included	Included	Included
Observations	890	615	890	615

Dependent Variable: pl = ln(EPO Filings/worker)

Standard errors in parentheses: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$

Notes pl denotes natural log of patents per worker, *L* one-period lag, *L*2 two-period lag, u = cycle/trend of ratio of real GDP and uR = cycle/trend ratio of Real Business R&D per worker, year denotes the time trend

Separate intercepts for each country (or country fixed effects) were included but the estimates are not reported to conserve space

cyclical movements in R&D do not seem to affect patent filings at the EPO. This suggests that patentable innovations are a function of long run R&D programs and are not influenced by short run boosts or declines in R&D funding. But these results are for all countries and sectors pooled. It remains to be seen whether the EPO filings of individual source countries or individual sectors are more sensitive to short run movements in R&D.

A third issue is whether different technological fields are affected differently by business cycles. Table 7 offers a preliminary look using the *International Patent Classification* (IPC) sectors. According to the results shown, EPO filings in Human Necessities (IPC A), Physics (IPC G), and Electricity (IPC H) are mildly procyclical (i.e., the coefficient of 'u' is statistically significant at the 10 % level), but filings in Chemistry (IPC C) are strongly procyclical (where the positive coefficient of 'u' is significant at the 5 % level). For all the IPC sectors except textiles (IPC D), EPO filings are strongly influenced by trend GDP. Still, the eight IPC sectors A to H are nonetheless very broadly defined. Each of these sectors consists of diverse fields of technology. Thus, future research could study the effects of cyclical shocks at a more disaggregated IPC level.

Another related issue is whether business cycles affect the EPO filings of different source countries differently. For good country-by-country analyses, more time-series observations are needed. Tentatively, therefore, this section provides a sample of results for the U.S., a major contributor to EPO filings. As shown in Table 8, business cycles have a positive and statistically significant effect on total U.S. filings at the EPO, where by 'total' it is meant aggregate sectors (see column 1). By individual sector, U.S. EPO filings in Performing Operations and Transportation (IPC B), Mechanical Engineering, Physics, and Electricity (IPC F, G, and H) are all significantly pro-cyclical. These results are for the annual sample. The quarterly sample results are quite similar (results not shown); the only difference is that the filings in Human Necessities are also significantly pro-cyclical in the quarterly sample. For EPO applications, the business cycle variable 'u' is only statistically significant for Electricity (IPC H), and only at the 10 % level (the results are also not shown to conserve space). With more data, a richer analysis of business cycles at the national level can and should be pursued.[10]

5   Other Applications

A side benefit of the modelling approach taken here is that the common effects over source countries for EPO filings can be analysed. Table 3 looked at correlations between countries in terms of their estimated intercept parameters. Another way is

---

[10]For U.S. domestic patenting, the longer time-series data of the United States Patent and Trademark Office (USPTO) could be exploited. For other countries, WIPO has very long time-series data on resident patenting. Lessons could be learned from these datasets as well.

Table 7 Extensions: business cycles by IPC sector

IPC code	Human necessities	Perform oper; transportation	Chemistry	Textiles	Fixed constructions	Mechanical	Physics	Electricity
	A	B	C	D	E	F	G	H
L.pl	0.609***	0.530***	0.621***	0.418***	0.397***	0.516***	0.667***	0.576***
	(0.032)	(0.032)	(0.033)	(0.038)	(0.035)	(0.032)	(0.033)	(0.033)
L2.pl	0.159***	0.188***	0.145***	0.239***	0.173***	0.147***	0.067**	0.146***
	(0.028)	(0.027)	(0.030)	(0.033)	(0.029)	(0.028)	(0.029)	(0.031)
yl	0.044***	0.450***	0.502***	0.406	0.543***	0.804***	0.625***	0.658***
	(0.094)	(0.097)	(0.120)	(0.269)	(0.137)	(0.123)	(0.124)	(0.134)
rl	0.018	0.088**	0.100**	0.203**	0.196***	0.062	0.152***	0.263***
	(0.042)	(0.042)	(0.050)	(0.083)	(0.052)	(0.051)	(0.048)	(0.060)
u	1.144*	0.204	1.674**	−0.325	−0.219	0.676	1.177*	1.462*
	(0.654)	(0.644)	(0.744)	(1.146)	(0.804)	(0.758)	(0.725)	(0.841)
year	0.007***	0.004*	0.000	−0.006	0.010***	0.005**	0.004*	0.004
	(0.002)	(0.002)	(0.002)	(0.004)	(0.003)	(0.002)	(0.002)	(0.003)
Country fixed effects	Included	Included	Included	Included	Included	Included	Included	Included
Observations	855	859	821	693	791	827	817	815

Dependent Variable: pl = ln(EPO Filings/worker)

Standard errors in parentheses: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$

Notes pl denotes natural log of patents per worker, L one-period lag, $L2$ two-period lag yl denotes real GDP per worker, rl real business enterprise R&D (BERD) per worker, and u the business cycle measure, where u = cycle/trend of GDP, and year denotes the time trend

Separate intercepts for each country (or country fixed effects) were included, but the estimates are not reported to conserve space

Table 8 Extensions: EPO Filings by the U.S. and Cyclical Shocks

Sector (IPC)	Overal aggregate	Human necessities	Perform Oper; transportation	Chemistry	Textiles	Fixed constructions	Mechanical	Physics	Electricity
	A		B	C	D	E	F	G	H
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$L.pl$	0.759***	0.610*	0.645**	0.409	0.452	0.239	−0.001	0.667***	0.847***
	(0.193)	(0.297)	(0.252)	(0.277)	(0.344)	(0.220)	(0.149)	(0.033)	(0.214)
$L2.pl$	−0.424**	0.008	−0.346	0.064	−0.060	−0.063	0.069	0.067**	−0.195
	(0.197)	(0.287)	(0.248)	(0.250)	(0.251)	(0.221)	(0.144)	(0.029)	(0.220)
rl	0.314	−0.468	−0.233	2.006**	−0.208	1.783**	0.305	0.625***	0.516
	(0.389)	(0.432)	(0.491)	(0.700)	(1.128)	(0.821)	(0.341)	(0.124)	(0.704)
yl	2.512***	2.438**	3.395***	0.442	3.162	−1.313	1.154**	1.154**	1.304
	(0.764)	(0.872)	(1.141)	(0.888)	(2.779)	(1.019)	(0.480)	(0.480)	(1.350)
u	2.201**	1.404	2.267**	0.576	1.718	1.705	2.522***	3.696**	4.372***
	(0.772)	(1.042)	(0.987)	(1.063)	(2.311)	(1.543)	(0.689)	(1.424)	(1.453)
year	−0.016**	−0.011	−0.030**	−0.041***	−0.053	0.037**	0.021**	−0.011	−0.015
	(0.007)	(0.009)	(0.012)	(0.013)	(0.042)	(0.017)	(0.009)	(0.012)	(0.013)
Constant	−3.453	−5.488	15.809	56.808***	64.464	−83.157**	−67.772***	−14.548	8.204
	(11.740)	(17.284)	(15.258)	(17.277)	(53.035)	(35.474)	(20.169)	(20.278)	(16.991)
Observations	22	22	22	22	22	22	22	22	22

Dependent Variable: pl = ln(EPO Filings/worker)

Standard errors in parentheses: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$

Notes pl denotes natural log of patents per worker, L one-period lag, $L2$ two-period lag

u = cycle/trend of ratio of real GDP and uR = cycle/trend ratio of Real Business R&D per worker

year denotes the time trend

Table 9 Correlation Coefficients between countries for one-year ahead 'transformed' filings

Model in differences				
Parameter	CN-HK	DE	JP	US
CN-HK	1	0.019	−0.006	0.020
DE	0.019	1	−0.022	0.015
JP	−0.006	−0.022	1	−0.006
US	0.020	0.015	−0.006	1

Notes From the model in first-differences for EPO filings, using the training data set from 1991 to 2013. The correlation coefficients between countries are for the one-year-ahead filings estimates on the transformed scale. Inter-country correlations are shown for China, Germany, Japan and US

to determine the correlation matrix of transformed filings estimates between countries ($\hat{\Xi}$ in Sect. 2). This may indicate common behavioural patterns between applicants in groups of countries that could be associated with the overall degree of industrialisation, possession of common industries with similar patenting behaviours, geographic proximity, and so forth. Table 9 shows the derived country correlation matrix for 2013 between the major source countries for the model in differences.

The correlations between these forecasts for total filings in 2013 from the most important countries are low and can be considered to be non-significant. From the full 28 country set, pairs of countries that have the largest positive correlations are those of Ireland with Greece (0.29) and with Spain (0.21); Portugal with Singapore (0.15); and China with Denmark (0.14). Pairs of countries with negative correlations are less prominent, the largest in size being Greece with Other countries (−0.13 − 0.09). While it may be tempting to suggest causal inferences about the relationships between these countries, one should be cautious about the power of the analysis and the lack of clarity regarding possibilities for common external causes for the associations. At any rate, such interesting questions about determining communalities of patenting behaviour between countries of origin go beyond the narrow requirements of forecasting for EPO's budgetary purposes.

6 Concluding Thoughts

The usefulness of the selected type of models has been demonstrated on historical EPO total filings data, and we recommend applying the model to year-to-year differences rather than to levels. We caution that the apparent widths of the 95 % confidence limits in diagrams such as Fig. 1 are somewhat too low, especially for the later parts of the forecasted horizon. In general, the quality of the forecasts

remains dependent on the assumed future levels of independent variables for GDP and R&D, the uncertainties of which are not explicitly considered in the confidence intervals.

For future research, it would be useful to further develop the model at the industry level or at the individual country level, as more observations become available. It would also be interesting to examine how business cycles in one country are transmitted internationally to affect innovation activity in another country. Specifically, the R&D and patenting of non-European countries may be affected not only by business cycles occurring in their own countries but also by fluctuations occurring in Europe. Other directions are more technically-oriented; for example, to study and incorporate the expectations formation of patent filers. EPO filings and applications may be driven by their expectations of future business cycles. In the models studied here, the contemporary value of 'u' entered the regression models. It may be the expected 'u', say u^e (the expected value of u in some future period), that affects current patenting decisions. Likewise, applicants may be affected by their forecasts of future trend GDP or future R&D budgets in designing their current innovation strategies. Modelling the expectations process is rather complex but should be worth exploring.[11] Another technical approach is to apply dynamic panel data methods, as in Hingley and Park (2015b), to account for the lagged dependent variables and country fixed effects.

Insofar as the aim of such research is to improve filings forecasts for the EPO budget, any method must take account of all the filings even where breakdowns of the data are considered separately or in parallel. Wider research into patenting behaviour can relax this requirement to some degree. The approach described here may turn out to be useful for such studies, including extensions to separate out particular patenting sectors or regions.

Finally, we believe that the way that the lognormal data were analysed here may be of some interest for wider areas of application. While the central limit theorem implies that lognormal data can indeed be described approximately in terms of normal error theory and the usual normal confidence limits, either care should be taken that the data sets are large enough for this or otherwise an approach such as the one described here should be used.

---

[11]See Dannegger and Hingley (2013) for a study on the use patent applicant surveys to infer the future expectations of applicants. In these surveys, applicants indicate their plans to file patents in various offices during the current year and the next two years. Their current intentions should reveal some information about their expectations of future market conditions.

References

Dannegger F, Peter H (2013) Predictive accuracy of survey-based forecasts for numbers of filings at the European Patent Office. World Patent Inf 35:189–200

Dikta G (2006) Time series methods to forecast patent filings. In: Hingley P, Nicolas M (eds) Forecasting innovations. Springer, Heidelberg

Draper N, Smith H (1981) Applied regression analysis, 2nd edn. Wiley, New York

European Patent Office (2015) Patent Filings Survey 2013: intentions of applicants regarding patent applications at the European Patent Office and other offices, 2014. http://www.epo.org/service-support/contact-us/surveys/future-filings.html

Hidalgo A, Gabaly S (2012) Use of prediction methods for patent and trademark applications in Spain. World Patent Inf 34(1):19–35

Hingley P, Nicolas M (2004) Methods for forecasting numbers of patent applications at the European Patent Office. World Patent Inf 26(3):191–204

Hingley P, Nicolas M (2006) Improving forecasting methods at the European Patent Office. In: Hingley P, Nicolas M (eds) Forecasting innovations. Springer, Heidelberg

Hingley P, Park W (2015a) A dynamic log-linear regression to forecast numbers of future filings at the European Patent Office. World Patent Inf 42(3):19–27

Hingley P, Park W (2015b) Do business cycles affect patenting? Evidence from European patent filings, Working Paper

Hodrick R, Prescott E (2007) Postwar business cycles: an empirical investigation. J Money Credit Banking 29:1–16

Johnson N, Kotz S, Balakrishnan N (1994) Continuous univariate distributions, vol 1. Wiley, New York

Meade N (2006) An assessment of the comparative accuracy of time series forecasts of patent filings: the benefits of disaggregation in space and time. In: Hingley P, Nicolas M (eds) Forecasting innovations. Springer, Heidelberg

Organisation for Economic Cooperation and Development (2014) Main science and technology indicators. http://www.oecd.org/sti/msti.htm

Park W (2006) International patenting at the European Patent Office: aggregate, sectoral and family filings. In: Hingley P, Nicolas M (eds) Forecasting innovations. Springer, Heidelberg

Soderlind P (2013) Lecture notes-econometrics: some statistics. http://home.datacomm.ch/paulsoderlind/Courses/OldCourses/EcmXSta.pdf