

Implementing a Forecasting Methodology for PCT Applications at WIPO

September 8th, 2003

DEHON Catherine[∇], ULB, ECARES

VAN POTTELSBERGHE Bruno^α, ULB, Solvay Business School

Abstract:

This paper investigates the effectiveness of several methods intending to forecast the number of yearly PCT applications at WIPO. Forecasting exercises have been applied for total PCT applications and for 5 countries accounting for more than 70 per cent of total PCT applications. So far, with the available data, the best ‘fit’ is obtained either with yearly data on total PCT and the AR(1) method (as opposed to country-specific estimations that have been subsequently aggregated for total PCT previsions) or with panel data estimates that include economic variables (GDP and R&D) for 5 countries. The forecasts of total PCT applications in 2002 range between 120 and 127 thousands units and between 140 and 150 thousands units in 2003. Several avenues for improvement are suggested, including an improved linearization of the basic series (other than logarithmic transformation), the use of sector specific data (as opposed to country-specific), and the use of national priority applications for the prevision of the forthcoming declining growth period (or ‘stationary’ period).

[∇] ECARES, Faculté SOCO, Institut de Statistique, Université Libre de Bruxelles (ULB), av F.D. Roosevelt 50 CP 114, 1050 Brussels, BELGIUM. Tel: +32-2-650.3858, E-mail: cdehon@ulb.ac.be.

^α Solvay Business School, Université Libre de Bruxelles (ULB), Solvay SA Chair of Innovation, Centre E. Bernheim, av F.D. Roosevelt 50 CP 145/1, 1050 Brussels, BELGIUM. Tel: +32-2-650.48.99, E-mail: bruno.vanpottelsberghe@ulb.ac.be. This research was partly performed when Bruno van Pottelsberghe was visiting professor at the Institute of Innovation Research (IIR), Hitotsubashi University, Tokyo, from July 2003 to December 2003.

1. Introduction

Since the start of the Patent Cooperation Treaty an increasing number of priority applications have gone through the PCT Process. PCT application have been booming for about twenty years now, witnessing the usefulness of allowing applicants to wait up to three years to decide whether it is worth it to enter into the international phase of protecting their inventions. For the Treaty itself the boom is a great success, but it probably creates some organisational complexities for WIPO authorities as yearly PCT applications jumped from about 5000 in the early eighties to 20.000 in the early nineties and well over 100.000 in the early 2000's.

It is well known that the statistical property of this kind of “non-stationary” series makes forecasting exercises more difficult to implement. The objective of this paper is to perform several methods intending to forecast the number of PCT applications at WIPO. Forecasting exercises are applied for total PCT applications and for 5 main countries accounting for about 80 per cent of total PCT applications. The focus is essentially put on the necessary steps required to implement an effective forecasting methodology.

Table 1. Potential forecasting methods of total PCT applications.

	PCT Series only		Economic Model	
	Yearly	Monthly	GDP, RD Yearly	GDP, RD and TO
Total PCT	√	√	√	√
Country	√	√	√	√

Table 1 presents the alternative methodologies that are used in this paper to forecast the number of PCT applications. Beside the statistical methods that are to be tested, several choices have to be done to test the validity of the forecasting techniques. For instance, one can focus exclusively on the available statistical series of PCT applications, or rely on an economic model that would take into account some economic variables (such as GDP, R&D expenses). This economic model can also be improved with some indicators of technological opportunity (TO) within each country. Furthermore, one can work with yearly data, quarterly data, or monthly data.

In what follows the forecasting “performance” of these methods is evaluated for both a short term prevision (1 year) and a medium term prevision (2 to 4 years).

The next section presents the broad statistical properties of the PCT time series and shows that they are far from stationary. We then apply a linearization process and perform two main forecasting methods (AR(1) and trend). The tests are performed on the total PCT application yearly and on 6 individual countries (that are subsequently added to get the global view). The ‘MAPE’ method is used to assess the forecasting quality of these methods. Section 3 reproduces a similar approach but with monthly data. The economic models are estimated in Section 4. The parameters estimated for these models are then used to implement additional forecast. Section 5 is devoted to a summary of the empirical findings, including actual forecast of PCT applications for the coming years) and a discussion on the potential improvements of forecasting techniques, in terms of raw data needs and statistical methodologies.

2. Yearly time series analysis of PCT applications

Figure 1 shows the annual number of PCT applications since 1978. It also gives the annual number of PCT for 6 countries and EPO priorities accounting for more than 70% of the total until 1979 and more than 80% since 1980.

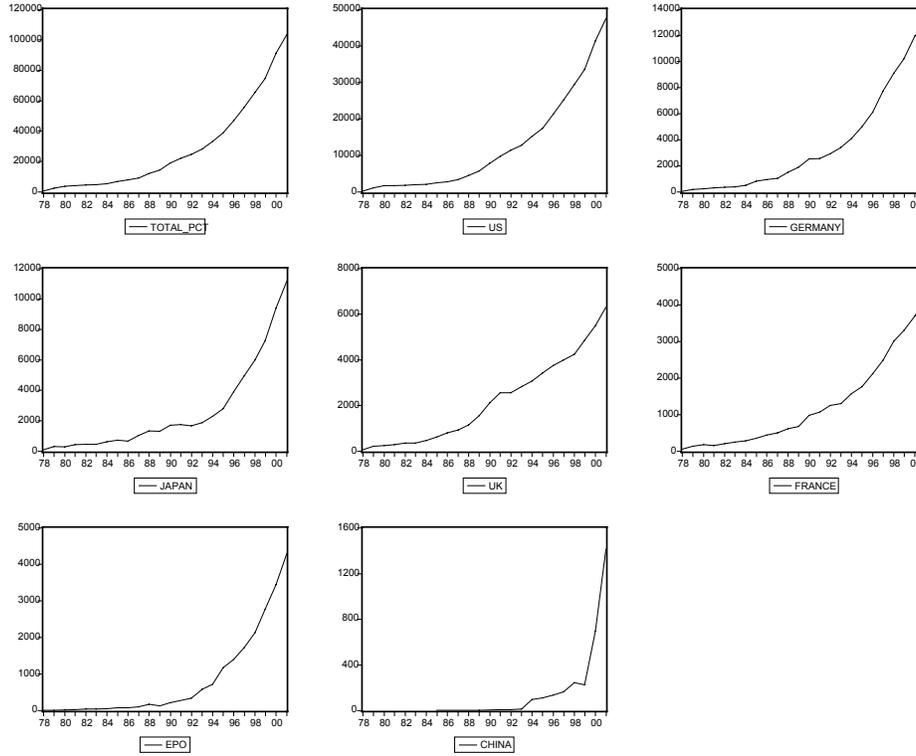
The first two years (1978-1979) witness the very beginning of the series and an early adaptation phase. They have therefore been dropped for the empirical exercise. The series that are used for all the forecasting methods start in 1980 and end in 2001. Figure 1 clearly shows that they follow an exponential form of the following type:

$$PCT_t = \alpha \beta^t + \text{error} \quad (1)$$

where α and β are the unknown parameters. In this equation, β represents the growth rate of PCT applications. To get a linear form of this relation, we take the logarithmic transformation (see Figure 2), hence we obtain a so called trend stationary process with a deterministic linear trend¹:

¹ A possible extension is to maximize the quality of forecast using other Box-Cox transformations.

Figure 1. Annual number of PCT since 1978 (total and selected countries)



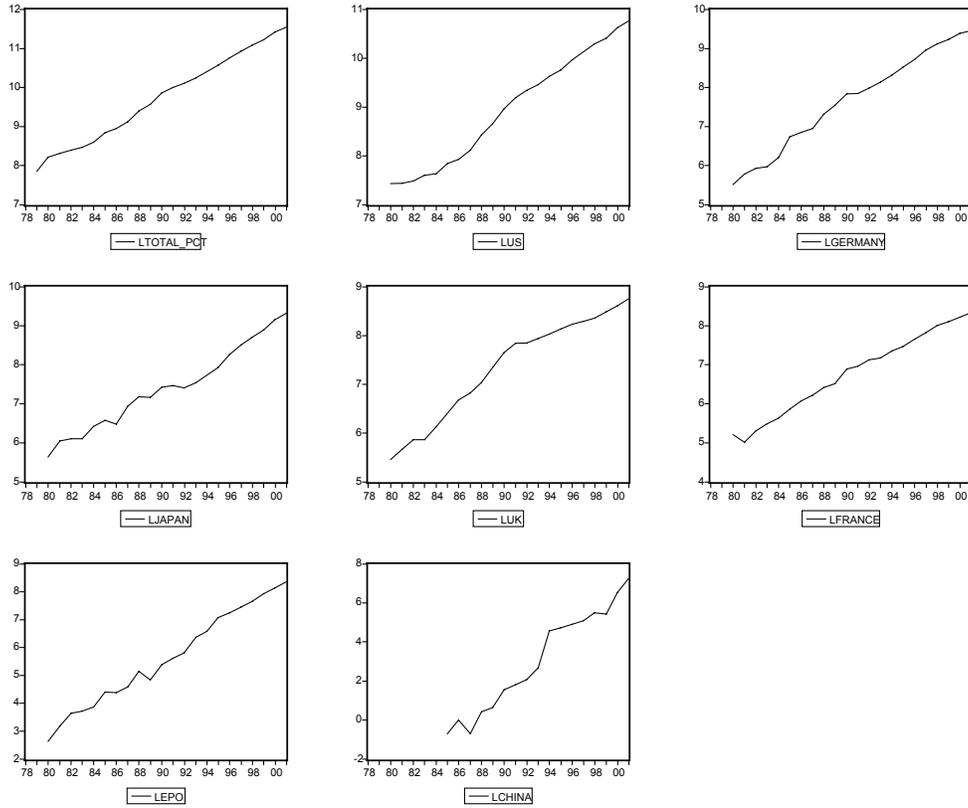
$$\log(\text{PCT}_t) = \log(\alpha) + \log(\beta) t + \varepsilon_t \quad (2)$$

where ε_t is the error term. Thanks to the well-known principle of decomposition of an additive model (with trend, without seasonality, without cycle) it is possible to implement a forecast using estimated trend.

Another more flexible direction is to exploit the Box-Jenkins method which consists in modelling the series to make them stationary, to chose an appropriate model and validate the model after estimation. The class of models used are the autoregressive integrated moving averages or ARIMA processes. These processes are the classical stationary ARMA processes after applying the first difference to obtain a stationary serie. The first step consists in obtaining series which are stationary, it is to say, series with mean, variance and covariance remaining constant over time. To reach this objective, the logarithmic transformation stabilizes the variance and the first difference stationarizes the mean (see Figure 3):

$$\Delta \text{PCT}_t = \log(\text{PCT}_t) - \log(\text{PCT}_{t-1}) \quad (3)$$

Figure 2. Number of PCT after logarithm transformation.



The transformed series ΔPCT_t is an approximation of the proportional growth rate (good approximation if the change in PCT is relatively small):

$$\Delta PCT_t \approx \frac{PCT_t - PCT_{t-1}}{PCT_{t-1}}. \quad (4)$$

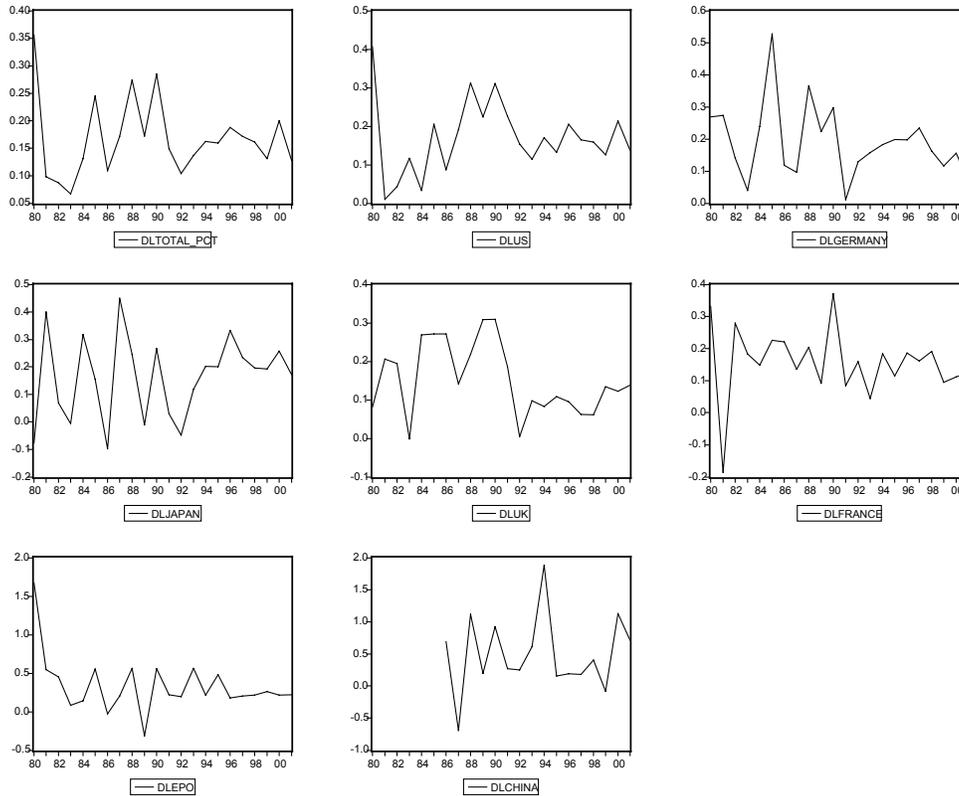
A model has to be specified for the transformed series. A usual model is the first order autoregressive one (AR(1)):

$$\Delta PCT_t = \mu + \gamma \Delta PCT_{t-1} + v_t \quad (5)$$

where μ and γ are the unknown parameters and v_t the error term. This model seems to be a good choice for the total number of PCT. If we split the series into several country series, the model can be improved. The problem in this case is that the analysis is more complex since we use different models for each country. Performing

forecasts on another model than the first order autoregressive model give worst results. We have for example estimated a moving average of order 2 (MA(2)) on the transformed US series, a MA(1) for the China and EPO transformed series, an ARMA(2,2) for Japan and Germany transformed series, an ARMA(1,2) for the UK and an ARMA(1,3) for France.

Figure 3. Number of PCT after mean and variance stationarisation.



Two methods to implement the forecast have been used to begin with:

- First, the additive model with trend (equation 2)
- Second, the first order autoregressive model (equation 5).

To observe the predictive power of each method, some tools are available: root mean square error (RMSE), mean absolute value (MAE), mean absolute percentage error (MAPE), etc. We use in the tables of results the third measure which is scale-invariant (it is not the case for the RMSE and MAE which depend on the scale of the dependent variable). The MAPE is given by:

$$MAPE = \frac{1}{h+1} \sum_s^{s+h} \left| \frac{\hat{PCT}_t - PCT_t}{PCT_t} \right| \quad (6)$$

where h is the number of periods for the forecast and \hat{PCT}_t is the forecast for time t. Measures on the variability of the forecast could also be considered.

To implement the forecast the main sample has been split in two independent parts: the “estimation” sample is used to perform the estimations and the “training” sample is used to compute the forecast. For instance, the estimates are run on the sample 1980-1997 and the 3 years forecast is computed for the period 1998 to 2001. The MAPE measures are reported in Table 2 for the forecast of 2, 3 and 4 periods using all the data since 1980.

We also calculate a weighted MAPE using the 7 countries in separate equations to estimate the total number of PCT (see Total* in table 2):

$$MAPE^* = \frac{1}{h+1} \sum_s^{s+h} \left| \frac{\sum_{i=1}^7 \hat{PCT}_{it} - \sum_{i=1}^7 PCT_{it}}{\sum_{i=1}^7 PCT_{it}} \right| \quad (7)$$

Table 2. MAPE measures for 2 forecasting methods and 2 to 4 years horizons.¹

Areas	2 years		3 years		4 years	
	Trend	AR(1)	Trend	AR(1)	Trend	AR(1)
Total	<u>2,11</u>	2,85	3,29	<u>2,16</u>	2,54	<u>1,62</u>
Total*	2.34	<u>1.19</u>	3.57	<u>1.95</u>	2.99	<u>1.92</u>
US	<u>2,89</u>	4,63	4,62	<u>1,94</u>	3,7	<u>1,5</u>
Germany	26,56	<u>11,56</u>	26,13	<u>18,1</u>	23,89	<u>18,19</u>
Japan	24,53	<u>10,39</u>	24,92	<u>10,2</u>	26,25	<u>12,73</u>
UK	45,33	<u>3,34</u>	53,87	<u>1,73</u>	65,56	<u>8,95</u>
France	16,67	<u>11,43</u>	16,2	<u>12,9</u>	13,69	<u>7,22</u>
EPO	<u>7,24</u>	10,00	<u>4,9</u>	9,36	<u>3,91</u>	16,34
China	<u>9,53</u>	53,04	59,3	<u>38,35</u>	68,67	<u>44,12</u>

1. MAPE measures for previsions on 2, 3 or 4 years using a training sample for the trend method and the AR(1) model. MAPE measures the average quality of forecasts on at least two years.

In the case of the total number of PCT series, the best forecast model is the AR(1) model for medium- to long term forecast (except for the two-years forecast of total PCT series, where the trend stationary model gives slightly better results). It must be

noticed, however, that the differences between the two statistical models is relatively small for both total PCT series and aggregate total (Total*) series. The USA and China series have the same behavior than the total number of PCT. For the other country series the picture is less clear. The AR(1) yields in all situations the smallest MAPE for Germany, Japan, the UK and France. On the other hand, the EPO series, where a MA(1) model is better to use, have better results with the trend stationary model. Concerning the value of the quality measure, China is the most difficult to forecast with huge MAPE, and the US and total series are associated with the smallest values.

The line Total* in Table 2 presents the MAPE* measure computed as in equation (7). It appears that the weighted sum of country specific forecast is slightly more efficient for medium term forecast (2 to 3 years) than implementing a forecasting model on total PCT applications. It is slightly less efficient for long term forecasting (4 years). Splitting the number of PCT series into additional country series might probably improve the quality of forecasting.

Table 3 presents the forecasts associated with the model using the total number of PCT (first line of Table 2). Again, it can be seen that for medium- to long-term (2 to 4 years) forecasts the predicted values based on the first order autoregressive model are better than those based on the trend stationary model. But for very short-term forecast (1 or 2 years) none of the models dominates.

Table 3. Forecast of the total number of PCT Applications based on total series.¹

Date	Data	4 years		3 years		2 years		1 year	
		Trend	AR(1)	Trend	AR(1)	Trend	AR(1)	Trend	AR(1)
1998	65468	65579	<u>65446</u>						
1999	74638	77644	<u>76719</u>	77616	<u>76746</u>				
2000	91123	91928	89533	91893	<u>89566</u>	91177	86974		
2001	103581	108841	<u>105563</u>	108797	<u>105604</u>	107888	<u>102400</u>	107876	<u>107532</u>

1. Forecast of the total number of PCT using the trend method and the AR(1) model on total yearly PCT applications.

Table 3b presents the forecast of total PCT applications based on country-level previsions (for the 5 largest applicants) that have been aggregated. Again, the AR(1) model performs better than the trend model, for both long term and short term

previsions. Comparing table 3 (AR(1) forecast on total PCT applications) with table 3b (aggregate forecast on 5 countries with AR(1) model), it seems that the latter approach yields better fits for one or two years forecasts, but the difference is not substantial.

Table 3b. Forecast of the total number of PCT Applications based on 5 countries' series.¹

Date	Data	4 years		3 years		2 years		1 year	
		Trend	AR(1)	Trend	AR(1)	Trend	AR(1)	Trend	AR(1)
1998	65468	68277	<u>65737</u>						
1999	74638	81672	<u>77348</u>	80780	<u>76982</u>				
2000	91123	98059	<u>91770</u>	96891	<u>91297</u>	95206	<u>88307</u>		
2001	103581	117055	<u>108284</u>	115544	<u>107666</u>	113376	<u>103897</u>	112208	<u>107263</u>

1. Forecast of the total number of PCT applications using the trend method and the AR(1) model on the yearly PCT applications of 5 countries (France, Japan, Germany, the UK, USA).

The next two sections contain two possible extensions for potential improvements of the forecasting performances. The first one consists in using the same models, but with monthly data. The second one relies on an economic model on annual data using more information on the 5 main countries included in this analysis.

3. Monthly time series analysis of PCT applications

The issue is now to test whether it would be possible to improve the ‘yearly’ forecast with monthly data. In this case the length of available series is significantly larger. Since the series on the monthly number of total PCT has an exponential structure, as the annual data, the transformed logarithm series has a linear behavior in time. For the method relying on an additive model structure the series can now be decomposed in a trend component and a seasonal components (including one dummy for each month). For the method relying on the Box Jenkins method, the seasonality is captured with the use of seasonal auto-regressif and/or moving average terms in the structure of the ARMA model. Then the ARMA model for the stationary series will be replaced by the seasonal autoregressive moving average model SARMA. The forecasts for the months from January 1998 to December 2001 based on monthly data from January 1980 to December 1997 are presented in Figure 4. The plain line shows the true data, the dotted line is obtained by estimation of SARMA(5,0)(1,0) on the stationary serie:

$$(1 - \gamma_1 L - \gamma_2 L^2 - \gamma_3 L^3 - \gamma_4 L^4 - \gamma_5 L^5) (1 - \gamma_{12} L^{12}) \Delta PCT_t = \mu + v_t$$

where $L(\cdot)$ is the lag operator. The dashed line gives a forecast based on an additive model with a trend and a seasonality component.

Figure 4. Previsions using a trend/seasonnality method (dashed) and a SARMA model (dotted).

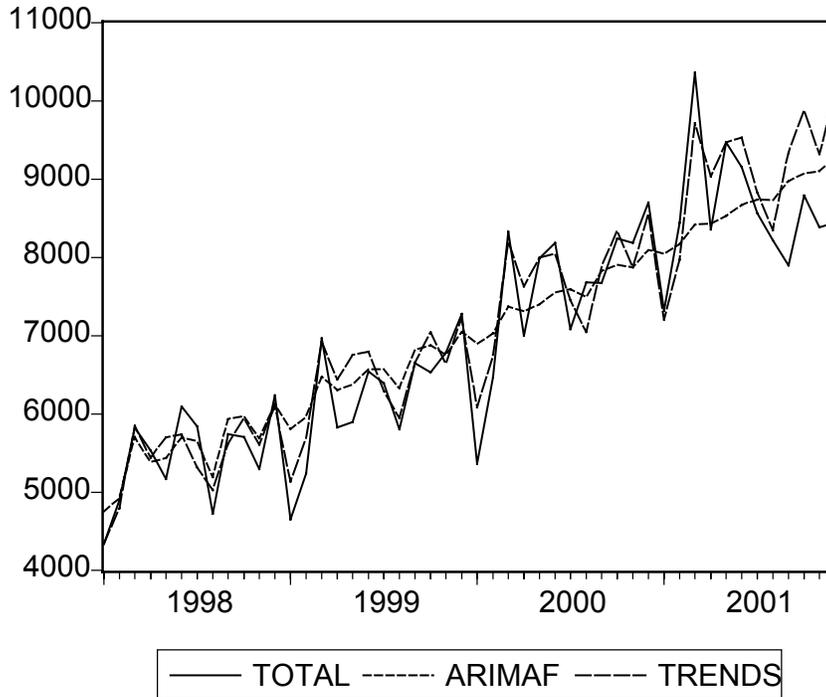


Table 4 presents a comparison of the AR(1) method on annual data and the two methods based on monthly data. For the sake of yearly comparison we have added the forecast of the 12 months. For short-term forecasting (1 to 2 years), the trend method on monthly data seems to be more efficient. It is very close to the ‘yearly’ forecasting for the one year forecast and outperform yearly forecast for the first year of two year forecast. Regarding the 4 years forecast, it yields either similar results (for the first two years) or better results for the third year. However, when compared with yearly forecasts, it does not seem that monthly evaluations on total PCT applications contribute to significant improvements of forecasting performances.

Table 4. Comparison of total PCT forecast using annual and monthly data.

		4 years			2 years			1 year		
		Year	Month.	Month.	Year	Month.	Month.	Year	Month.	Month.
Date	Data	AR(1)	ARIMA	Tr.+S	AR(1)	ARIMA	Tr.+S	AR(1)	ARIMA	Tr.+S
1998	65468	<u>65446</u>	66480	65492						
1999	74638	<u>76719</u>	77904	77559						
2000	91123	89533	90343	<u>91849</u>	86974	88765	<u>91048</u>			
2001	103581	105563	<u>104177</u>	108773	<u>102400</u>	103917	107756	<u>107532</u>	109375	107663

An alternative approach is to perform monthly forecasts on country series and aggregate these series afterwards (using the constant share of the five countries in total PCT applications). Table 4b shows that the forecast are better with an ARIMA model than with a trend model, but they never outperform the forecasts based on total monthly PCT applications series. It is not clear whether splitting the total series into additional country series would improve the results.

Table 4b. Total PCT forecast using monthly data on 5 country-series

		4 years		2 years		1 year	
		Trend+S	ARIMA	Trend+S	ARIMA	Trend+S	ARIMA
1998	65468	68315	<u>67726</u>				
1999	74638	81793	<u>80751</u>				
2000	91123	98339	<u>95853</u>	95320	<u>88970</u>		
2001	103581	117463	<u>112609</u>	113565	<u>104663</u>	112256	<u>108023</u>

1. Forecast of the total number of PCT applications using the trend seasonal method and the ARIMA models on the monthly PCT applications of 5 countries (France, Japan, Germany, the UK, USA).

4. Economic models

Another direction to improve the method using annual data can be to use additional information about the countries. In what follows this method is applied to the 5 countries representing more than 70% of total of PCT application (France, Germany, Japan, the UK and the United States). The extrapolation for the total number of PCT applications will be computed using the relative weight of the total of the 5 countries for the last available year.

The two main independent variables are the total domestic R&D outlays in 1995 constant US PPP's (millions) and the gross domestic product in 1995 constant US PPP's (millions).

To avoid the problem of spurious regression (due to the non stationarity of the processes), we use the transformed series of PCT, as in equation (3). The basic economic model is a panel data model using country-specific fixed effect terms (α_i):

$$\Delta PCT_{it} = \alpha_i + \delta \Delta PIB_{it-1} + \gamma \log(PIB_{it-1}) + \phi \Delta DIRD_{it-2} + \beta \log(DIRD_{it-1}) + \varepsilon_{it} \quad (8)$$

where the index $i=\{1,2,3,4,5\}$ indicates the countries and

$$\Delta PIB_{it} = \log(PIB_{it}) - \log(PIB_{it-1}),$$

$$\Delta DIRD_{it} = \log(DIRD_{it}) - \log(DIRD_{it-1}),$$

as in equation (3) for PCT applications. The logarithm transformation is used to stabilize the variance of the dependant variable and it further allows to interpret each estimated parameter as an elasticity. Equation (8) intends to explain the growth rate of PCT applications with the lagged GDP and R&D, both in level and in growth rates.² Growth rate variables reflect more short term impacts, whereas the levels variables reflect more long term equilibrium. For GDP a one year lag is used for both the level variable (simple logarithm transformation) and the growth rate variable. Regarding R&D, a one year lag is used for the level and a two years lag for the growth rate. It is assumed that the short term impact of additional R&D efforts take about two years to translate into a patent application.

Equation 8 can be complemented by an approximation of technological opportunity, which would reflect the extent to which new technologies develop fast in a given

² One alternative was to include the level of the number of PCT lagged one year ($\log PCT_{t-1}$) on the right-hand side of equation (8). We could also have added the one year lagged growth rate of PCT (ΔPCT). However, when the lagged dependent variable is included amongst the explanatory variables, the hypothesis of exogeneity does not hold anymore ($\text{cor}(\log PCT)_{t-1}, \varepsilon_t)=0$). The Breusch-Godfrey Lagrange multiplier test allows to measure the degree of serial correlation. For the case of France and the USA, the errors are not serially correlated but for the UK, Japan and Germany the hypothesis of exogeneity does not hold. To avoid this bias, it has been decided not to use the lagged dependent variable as explanatory variable. Without the lagged dependent variable the serial correlation is substantially lower, which suppresses the endogeneity issue.

country. Two main technological opportunity indicators are used; one for the ICT (information and communication technologies) and the other one for the biotech sector. The two variables are computed as follows:

BIOT = number of patents applications in biotech / total patent applications

ICT = number of patents applications in ICT / total patent applications

The data comes from the OECD MSTI database on patent applications at the European Patent Office (EPO) for the five countries. We can either assume that technological opportunity variables are part of the independent variables, as in equation (9), or that technological opportunity must be interacted with the R&D variable, as in equation (10).

$$\Delta PCT_{it} = \alpha_i + \delta \Delta PIB_{it-1} + \gamma \log(PIB_{it-1}) + \phi \Delta DIRD_{it-2} + \beta \log(DIRD_{it-1}) + \phi BIOT_{it-1} + \phi ICT_{it-1} + \varepsilon_{it} \quad (9)$$

$$\Delta PCT_{it} = \alpha_i + \delta \Delta PIB_{it-1} + \gamma \log(PIB_{it-1}) + \phi \Delta DIRD_{it-2} + \beta^c \log(DIRD_{it-1}) + \beta^\phi BIOT_{it-1} \log(DIRD_{it-1}) + \beta^\phi ICT_{it-1} \log(DIRD_{it-1}) + \varepsilon_{it} \quad (10)$$

In the latter case, the estimated equation would assume that the elasticity of PCT application with respect to R&D (β) is composed by a fixed component (β^c) and a component that depends on the two technological opportunity variables, as follows:

$$\beta = \beta^c + \beta^\phi BIOT + \beta^\phi ICT \quad (11)$$

In other words, equation (10) allows to test whether the impact of R&D on PCT applications varies with respect to the relative importance of two high-tech sectors in an economy, namely the ICT and bio-tech sectors.

The estimates of equations (8), (9) and (10) are presented in Table 5. The sample is composed of five major applicant countries for the period 1981-2000. It clearly appears that equation (8) does not perform very well, as the F-statistic (test for the joint significance of the estimated parameters) is not significant, the adjusted R-

squared is very low, and only one parameter is significantly different from zero. A low R-squared is a frequent fact with first differenced variables and is not a sign of “low model”. Introducing the technological opportunity variables slightly improves the performance of the model. The F-statistic of equation (9) is significant and the two variables of technological opportunity are also significantly different from zero. The countries that have a larger share of EPO patent in the bio-tech sector (i.e. who have relatively more inventions in that field) are associated with a higher growth of their PCT applications. The ICT sector seems to have an opposite effect, with the countries that have a high share of inventions in ICT being associated with a lower growth of their patents.

Table 5. Panel data estimates of PCT applications, 1981-2000.¹

Dependent var. is ΔPCT_{it}	Eq. (8)	Eq. (9)	Eq. (10)
log (PIB _{it-1})	-0.278 * <i>0.162</i>	-0.291 <i>0.225</i>	-0.322 <i>0.219</i>
ΔPIB_{it-1}	0.014 <i>0.553</i>	-0.245 <i>0.556</i>	-0.280 <i>0.555</i>
log (DIRD _{it-1})	0.187 <i>0.136</i>	0.301 <i>0.155</i>	0.340 * <i>0.153</i>
$\Delta DIRD_{it-2}$	-0.505 <i>0.391</i>	-0.439 <i>0.414</i>	-0.466 <i>0.415</i>
BIOT _{it-1}		2.372 * <i>1.174</i>	
ICT _{it-1}		-1.323 * <i>0.631</i>	
BIOT _{it-1} * log (DIRD _{it-1})			0.215 * <i>0.103</i>
ICT _{it-1} * log (DIRD _{it-1})			-0.111 * <i>0.057</i>
F-stat	1.865	2.677 *	2.676 *
Adjusted R-squared	0.002	0.027	0.033

1. Within estimates, all equations include country-specific dummies, standard errors are in *italic*; * indicates that the parameter is significant at the 10% probability threshold. The panel includes five countries (France, Germany, Japan, the UK, and the United States) over the period 1981-2000.

Equation (10) also takes the technological opportunity variables into account, but as an interaction with the level of R&D investments. The results are also better than for equation (8) and confirm to some extent the estimates of equation (9). Indeed, the countries with a high share of inventions devoted to the bio-tech sector benefit from a

higher elasticity of PCT applications with respect to R&D expenses. The reverse is true for the share of inventions devoted to the ICT sector.

An additional model has been run (not reported here for the sake of space) with country-specific parameters for the interaction between R&D outlays and the technological opportunity variables (eq. 10-c).

The estimated parameters of Table 5 have been used to implement several one-year forecasts of total PCT applications. The weighted sums of the country-specific forecasts are presented in Table 6. The forecast of the simpler model, equation (8), yields better results. On average, the absolute error fluctuates around 2.000 PCT applications. There is however a clear cyclical effect that is not corrected by the economic model. Indeed, all the forecasts are overestimated for the years 2001 and 1999; whereas they are underestimated for the year 2000.

Table 6. Forecast error (1 year) with the economic panel data models¹

year	Eq.(8)	Eq.(9)	Eq.(10)	Eq.(10-c)
2001	-1420	-1653	-1660	-3091
2000	4839	5842	5823	4808
1999	-856	-1436	-1623	-2578
1998	442	-595	-763	-468
Mean absolute value	1889,25	2381,5	2467,25	2736,25

1. One year aggregate forecast errors for the total PCT applications. Computation based on panel data estimates on 5 countries (France, Germany, Japan, the UK, and the United States), see Table 5.

5. Forecast intervals

The number of PCT applications (both total and for each country) is still non-stationary, which means that no stable trend has been reached yet. This non-stationarity is one of the main reasons why forecasting the number of PCT is far from being straightforward.

The choice of the model depends on the criterion that is used. One can either chose the model that provides the best forecast for the year 2001, or the model that provides the best one-year forecast over the past 4 years. Table 7 and table 7b show the

forecast errors for all the models that have been used in the previous sections. If the first criterion is used, the most accurate forecast model is in the row “2001”. In this case it seems that Equation 8 of the economic panel data model would provide the most accurate forecast, with country-specific estimations that have been subsequently aggregated for total PCT previsions. If the second criterion is used, the most accurate forecast model is to be identified in the row “mean absolute value”. Here, the best fit seems to be performed by time series analyses (yearly or monthly, with trend) of total PCT applications.

Table 7. Forecast errors of total PCT applications (1 year)

	Time series ¹				Panel data/economic data ²			
	Year	Year	Month	Month	Year	Year	Year	Year
	Trend	AR(1)	Trend	Arima	Eq.(8)	Eq.(9)	Eq.(10)	Eq.(10-c)
2001	-4295	-3951	-4254	-5966	-1420	-1653	-1660	-3091
2000	-54	4149	-136	2147	4839	5842	5823	4808
1999	-2978	-2108	-2950	-306	-856	-1436	-1623	-2578
1998	-111	22	-75	-1063	442	-595	-763	-468
Mean absolute value	1860	2558	1854	2371	1889	2382	2467	2736

1. One year forecasts errors of total PCT applications based on time series analysis of total PCT applications at WIPO.
2. One year forecasts errors of total PCT applications based on panel data estimates of 5 countries (France, Germany, Japan, the UK, and the United States).

Table 7b. Forecast errors of total PCT applications (1 year)

	Time series ³			
	Year	Year	Month	Month
	Trend	AR(1)	Trend	Arima
2001	-8627	-3682	-8847	-4614
2000	-4083	2816	-4408	1942
1999	-6142	-2344	-6271	-30
1998	-2809	-269	-2898	-2309
Mean absolute value	5415	2278	5606	2224

3. One year forecast errors of total PCT applications based on time series analysis of 5 country-series PCT applications (France, Germany, Japan, the UK, and the United States).

The forecast for the next 2 years (2002 and 2003) using the techniques presented above are presented in table 8 and table 8b. The actual forecasts are actually close to each other. If the first criterion is used (best forecast on the year 2001), the economic panel data method is more relevant, with forecasts of about 120000 PCT applications in 2002, and 141000 PCT applications in 2003.

If the second criterion is used (best average one year forecast over the past four years), the time series analyses of total PCT applications (with trend) provide the best fit. The actual forecast is higher than with the economic panel data model, with about 126500 PCT applications for 2002 and 149000 PCT applications for 2003.

Table 8. Total PCT forecasts (1000's) for the years 2002 and 2003

	Time series ¹				Panel data/economic data ²			
	Year	Year	Month	Month	Year	Year	Year	Year
	Trend	AR(1)	Trend	Arima	Eq.(8)	Eq.(9)	Eq.(10)	Eq.(10-c)
2002	126.7	120.7	126.5	107.3	120.0	122.0	122.2	124.0
2003	149.9	141.3	149.6	115.2	140.5	145.3	145.8	150.1

1. Total PCT applications forecast based on time series analysis of total PCT applications at WIPO.
2. Total PCT applications forecast based on panel data estimates of 5 countries (France, Germany, Japan, the UK, and the United States).

Table 8b. Total PCT forecasts (1000's) for the years 2002 and 2003

	Time series ³			
	Year	Year	Month	Month
	Trend	AR(1)	Trend	Arima
2002	131.7	124.7	131.8	111.4
2003	157.0	151.8	157.2	125.8

3. Total PCT applications forecast based on time series analysis of 5 country series PCT applications (France, Germany, Japan, the UK, and the United States).

Several avenues for improvement can be implemented, with respect to raw data availability, statistical methods, and the use of additional types of economic series:

- National priority applications (yearly and monthly, including an IPC classification at 2 Digits) would be used for benchmarking the countries in terms of the propensity to rely on the PCT process. These benchmarks would allow to perform a prevision of the forthcoming declining growth period (or 'stationary' period), that must happen in the coming years.
- Quarterly economic data (e.g. output) might even improve the the panel data econometric analysis further.
- The use of sector specific data (as opposed to country-specific), would lead to an identification of broad technological revolutions. We are convinced that for some sectors the series of PCT applications are already stationary.

- Using more countries (or group of countries) would probably improve the performance of time series analyses, both yearly and monthly.
- An improved linearization process of the basic series might also improve the statistical fit. We use a log-linear transformation. Simulations would induce the use of more precise linearization process and therefore would improve our forecasts.

Bruno van Pottelsberghe is Vice President of the Solvay Business School. He has been professor at the Brussels' University (ULB) since September 1999. As holder of the *Solvay S.A. Chair of Innovation* he teaches courses related to the economics and management of innovation and intellectual property. Regarding his main administrative duties, he is Vice-president of the Solvay Business School, Director of the MBA Programs and of the International Exchange Programme.

Before joining the teaching staff of the ULB, Bruno van Pottelsberghe worked two years at the OECD (Department of Science, Technology, and Industry), and several months as visiting researcher at the Columbia Business School (NYC) and at the Research Institute of the MITI (Tokyo). As a Ph.D. candidate in Economics, he was Research Fellow at the Department of Applied Economics of the ULB (DULBEA). He performed his Ph.D. ("*The effectiveness of S&T policies inside the Triad*") after a Bachelor's degree in Economics, a Master in International Relations and a Master in Econometrics (ULB). Bruno van Pottelsberghe is currently (July to December 2003) Visiting Professor at Hitotsubashi University, Institute of Innovation Research (IIR).

For further information : <http://www.ulb.ac.be/cours/solvay/vanpottelsberghe>

Selected publications related to patent data and S&T policies

"What patent data reveals about universities – The case of Belgium", jointly with S. Saragossi, *Journal of Technology Transfer*, 28(1), 2003, pp. 47-51.

"The value of patents and patenting strategies : countries and technology areas patterns", jointly with D. Guellec, *Economics of Innovation and New Technologies*, 11(2), 2002, pp. 133-148.

"The internationalisation of technology analysed with patent data", jointly with D. Guellec, *Research Policy*, 30(8), pp. 1256-1266, 2001.

"Applications, grants and the value of patents", jointly with D. Guellec, *Economic Letters*, 69(1), pp. 109-114, 2000.

"The impact of public R&D expenditure on business R&D", jointly with D. Guellec, *Economics of Innovation and New Technologies*, 2002.

"R&D and productivity growth – A panel data analysis of 16 OECD countries", (English and French) jointly with D. Guellec, *OECD Economic Studies*, 2001, pp. 103-126.

"Does foreign direct investment transfer technology across borders?", jointly with F. Lichtenberg, *Review of Economics and Statistics*, 83(3), 2001, pp. 490-497.

Catherine Dehon holds a Ph.D. in Statistics from the Université Libre de Bruxelles, 2001. She has contributed to the development of robust statistical methodology in regression problems and multivariate analysis. Her current research interest is the robustification of econometric methods. She is assistant professor at ULB where she teaches statistics and econometrics.

For further information: <http://student.ulb.ac.be/~cdehon/>

Selected publications

Dehon, C., Filzmoser, P. and Croux, C. (2000), "Robust Methods for Canonical Correlation Analysis", in *Data Analysis, Classification, and Related Methods*, eds H.A.L. Kiers, J.P. Rasson, P.J.F. Groenen, M. Schrader, Berlin : Springer-Verlag, 321--26.

Croux, C., Dehon, C., Rousseeuw, P.J., and Van Aelst S. (2001), "Robust Estimation of the Conditional Median Function at Elliptical Models", *Statistics and Probability Letters*, **51**, 361--368.

Croux, C., and Dehon, C. (2001), "Robust Linear Discriminant Analysis using S-estimators", *The Canadian Journal of Statistics*, **29**, 473--492.

Croux, C., and Dehon, C. (2002), "Analyse Canonique basée sur des Estimateurs Robustes de la Matrice de Covariance", *La Revue de Statistique Appliquée*, **L (2)**, 5-26.

Croux, C., and Dehon, C. (2002), "Estimators of the Multiple Correlation Coefficient: local robustness and confidence intervals", to appear in *Statistical Papers*.

Croux, C., Van Aelst, S., and Dehon, C. (2002), "Bounded Influence Regression using High Breakdown Scatter Matrices", to appear in *Annals of the Institute of Statistical Mathematics*.

Dehon, C., and Croux, C. (2002), "Statistical Inference for a Robust Measure of Multiple Correlation", in *Proceedings in Computational Statistics 2002*, eds W. Härdle, B. Rönz, Berlin : Physica-Verlag, 557—562.

Implementing a Forecasting Methodology for PCT Applications

Catherine DEHON, ULB

Bruno VAN POTTELSBERGHE, ULB and IIR

WIPO Conference, 18-19 September 2003

Today's Menu

- Objectives
- Series of PCT Applications
- Time Series Analysis
- Economic Model and Panel Data
- Concluding Remarks and Next Steps

Today's Menu

- ➔ • **Objectives**
- **Series of PCT Applications**
- **Time Series Analysis**
- **Economic Model and Panel Data**
- **Concluding Remarks and Next Steps**

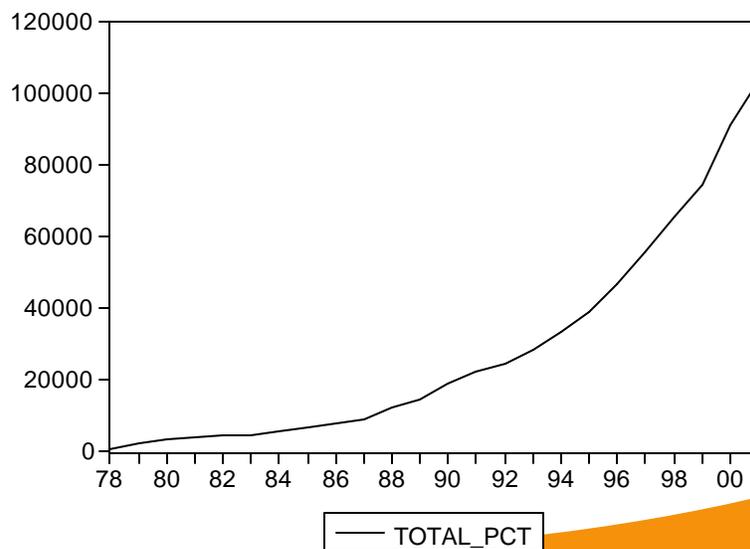
Objectives

- **To Implement a Relevant Forecasting Model of PCT Applications**
 - Analyse time series
 - Identify several potential methods
 - Test the effectiveness of these methods
 - To provide the forecasts for 2002 and 2003
 - Set recommendations for improved forecasts

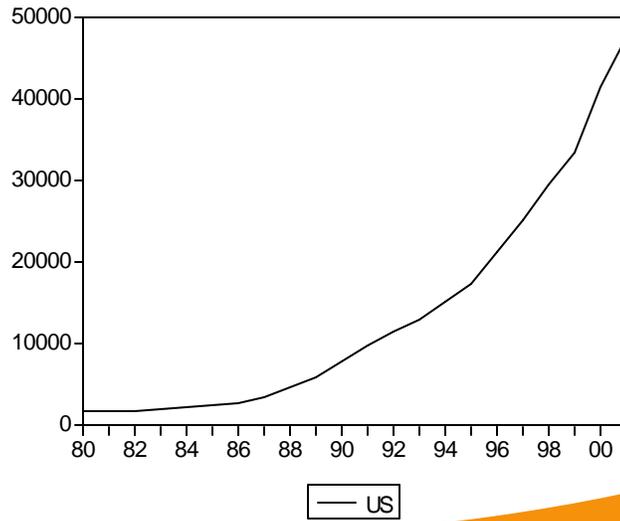
Today's Menu

- Objectives
- ➔ • Series of PCT Applications
- Time Series Analysis
- Economic Model and Panel Data
- Concluding Remarks and Next Steps

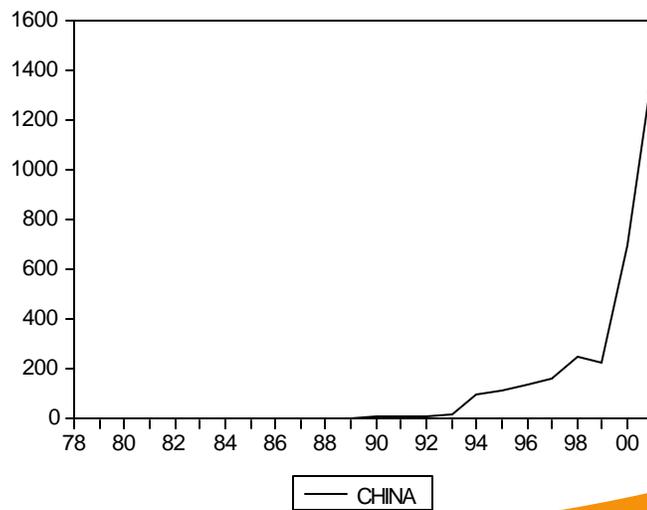
Series of PCT Applications



PCT Applications – USA (50.000)



PCT Applications – China (1.400)

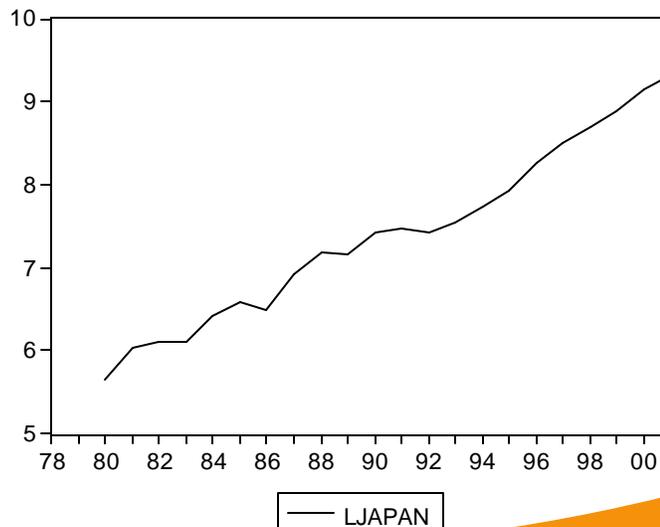


Series of PCT Applications

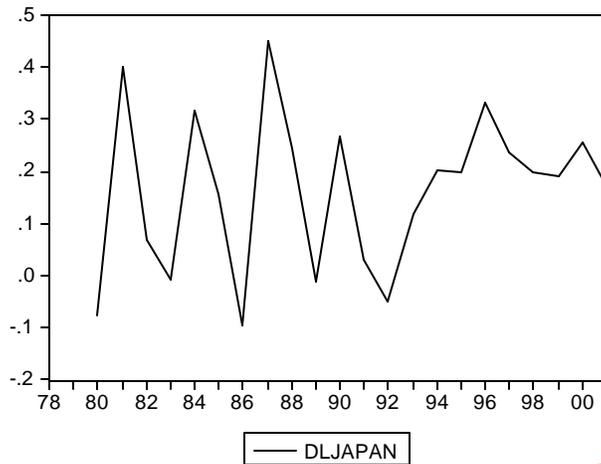
- Far from stationarity
- Logarithmic transformation
- First difference of logs

	PCT ~ 2001
Total	105.000
USA	50.000
Japan	11.000
Germany	12.000
France	> 4.000
China	1.400

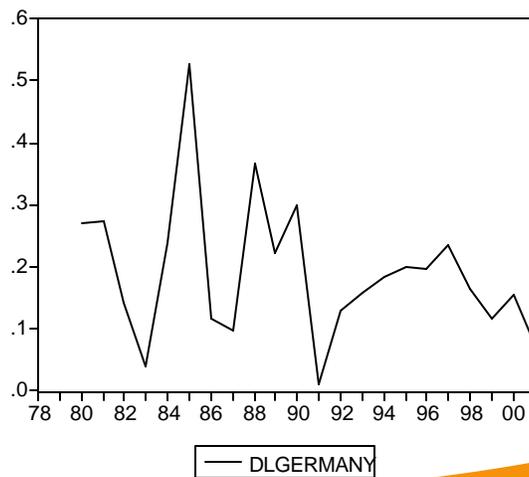
Log (PCT Applications) – Japan



Log Dif (PCT Applications) – Japan



Log Dif (PCT Applications) – Germany



Forecasting models :



Series Only:

- Yearly vs. Monthly
- Total PCT vs. Country series
- Trend (log) vs. AR1 (dlog)

Economic Panel data Model

- GDP, RD
- GDP, RD, TO

Today's Menu

- Objectives
- Series of PCT Applications
- ➔ • Time Series Analysis
- Economic Model and Panel Data
- Concluding Remarks and Next Steps

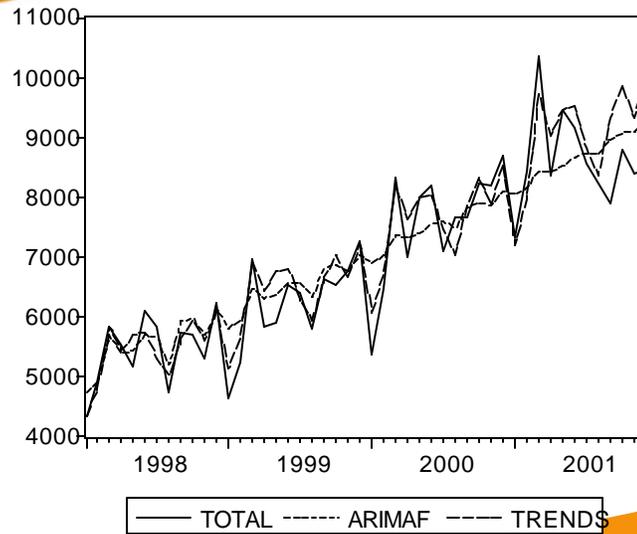
Table 2. MAPE measures for 2 forecasting methods on total PCT applications, and country series

Areas	2 years		3 years		4 years	
	Trend	AR(1)	Trend	AR(1)	Trend	AR(1)
Total	<u>2,11</u>	2,85	3,29	<u>2,16</u>	2,54	<u>1,62</u>
Total*	2.34	<u>1.19</u>	3.57	<u>1.95</u>	2.99	<u>1.92</u>

Table 3. Forecast of the total number of PCT Applications based on total series (000's)

Date	Data	4 years		2 years		1 year	
		Trend	AR(1)	Trend	AR(1)	Trend	AR(1)
1998	65.5	65.6	<u>65.4</u>				
1999	74.6	77.6	<u>76.7</u>				
2000	91.1	<u>91.9</u>	89.5	<u>91.2</u>	87.0		
2001	103.6	108.8	105.6	107.9	102.4	107.9	107.5

Figure 4. Monthly previsions using trend/seasonality and SARMA models



Conclusions for time series :

- AR(1) ~ Trend for total PCT short term
- AR(1) > Trend for country series
- Country series slightly better than total PCT
- Monthly series not better than yearly series
- Monthly/country series not better

Today's Menu

- Objectives
- Series of PCT Applications
- Time Series Analysis
- ➔ • Economic Model and Panel Data
- Concluding Remarks and Next Steps

Economic Model and Panel Data :

$$\Delta PCT_{it} = \alpha_i + \delta \Delta PIB_{it-1} + \gamma \log(PIB_{it-1}) \\ + \varphi \Delta DIRD_{it-2} + \beta \log(DIRD_{it-1}) + \varepsilon_{it}$$

And TO, both independent and interacted with DIRD, only significant elasticities (table 5), but no improvement of forecast errors (table 6).

Today's Menu

Objectives

Series of PCT Applications

Time Series Analysis

Economic Model and Panel

➔ Data

Concluding Remarks and

Next Steps

Table 7. Forecast errors of total PCT applications (1 year)

	Time-series ¹				Panel data/economic data ²			
	Year	Year	Month	Month	Year	Year	Year	Year
	Trend	AR(1)	Trend	Arima	Eq.(8)	Eq.(9)	Eq.(10)	Eq.(10-c)
2001	-4295	-3951	-4254	-5966	-1660	-3091		
2000	-54	4149	-136	2147	4839	5842	5823	4808
1999	-2978	-2108	-2950	-306	-856	-1436	-1623	-2578
1998	-111	22	-75	1063	442	-595	-763	-468
Mean absolute value	2558	2371	1889	2382	2467	2736		

**Table 8. Total PCT forecasts (1000's)
for the years 2002 and 2003**

Time series ¹				Panel data/economic data ²			
Year	Year	Month	Month	Year	Year	Year	Year
Trend	AR(1)	Trend	Arima	Eq.(8)	Eq.(9)	Eq.(10)	Eq.(10-c)
2002	120.7		107.3		122.0	122.2	124.0
2003	141.3		115.2		145.3	145.8	150.1

Avenues for improvements



- More country series
- Sectoral analysis
- Quarterly economic data
- National priority applications.... Kinked curve approximation
- Country specific linearization