Summary

Back to the future
The call for justice, 2008-2017: forecasts and actual figures

Good forecasts of the future need for capacity of all organizations in the field of justice are essential input for the government budget for justice expenses. For nearly two decades these forecasts have been made annually by the WODC in collaboration with the Ministry of Justice and Security using advanced statistical methods. It is assumed that the justice system more or less operates like a supply chain: a system of organizations, people, activities, information, and resources involved in moving a "justice problem" (i.e. a crime or a dispute) from its origin to its resolution. The justice problem may re-enter the supply chain at any point, for example when fines are not paid or agreements are not observed. We therefore speak of a justice chains.

The forecasts consist of two parts. The first part are the policy-neutral forecasts. These forecasts are made under the assumption that no changes in policy or legislation occur. They are made with a statistical model (PMJ), which relates economic, demographic, societal and institutional developments to crime trends and trends in civil and administrative suits by performing a regression analysis on time series. These estimated relationships do not necessarily imply causality: the focus is on the predictive power. Next, the supply chain idea is implemented: the PMJ-model relates the (forecasted) trends in crime and civil and administrative suits to developments further on in the justice system, such as the public prosecutor, the court of appeal and sentencing. Thus the PMJ-model becomes a justice chains model. The policy-neutral forecasts produced by this model are published annually by the WODC en the Council for the Judiciary.

For the second part of the forecasts expected effects of changes in policy or legislation are estimated. These are made by the Ministry of Justice and Security and its executive agencies. Various techniques are used. The estimated policy effects together with the policy-neutral forecasts form the policy-enriched forecasts. This report investigates the accuracy of the policy-enriched forecasts, the accuracy of the starting values on which the forecasts are based and the effect of adding policy effects on the forecast error over the past ten years.

Forecasts are made up to seven years ahead. The first year is the year in which the forecasts are calculated and the second year is the year in which the forecasts are published. Three to seven years ahead is the span of the government budget. To calculate the forecast error the forecasts are grouped by year ahead. The absolute percentage error is calculated for a set of key variables to avoid canceling out positive and negative values. Then the average is calculated over all key variables. The mean absolute percentage error for forecasting one year ahead (the year in which the calculations are made) is, on average, 8% and increases to 20% for three years ahead (first year of the government budget) and to 58% for seven years ahead (last year of the government budget). On average there is an overestimation, but it does not seem to be systematic. The model produces both forecasts that are too high as well as forecasts that are too low.

The equations in the statistical model are mainly formulated in terms of first differences of either the absolute values or the natural logarithms. As a result, the statistical model produces growth forecasts. The absolute values are calculated by applying the predicted growth to a starting value. This could be the last known annual value (usually the previous year) or a preliminary estimate of the annual value for the current year. An incorrect starting value can have a considerable effect on the forecasts of the absolute values. It turns out that, after the forecasts have been published, the last known annual values are actually adjusted downwards with an average of 3%, even though they were supposed to be final. Presumably, the values for earlier years are also adjusted, but this has not been verified. The average downward adjustment of the preliminary estimates of the annual value for the current year is even larger: 7%. So, in general, the starting point for the forecasted growth path is, in hindsight, too high. Thus, even if the growth would have been perfectly forecasted, the forecast error of the absolute forecasts would still be 7%.

The estimated effects of new legislation or policy generally have an upward effect on the forecast error. Without the policy effects the medium and long term forecast errors would have been 1 to 1½% lower. Because the medium and long term forecast errors are fairly large (up to 58% in the 7th year), this effect is limited. Nevertheless, the pace with which new policy of legislation comes into effect seems to be overestimated. It is also possible that the expected effect is in reality lower or non-existent or that new policies have already been effective before the official starting date. Notably, the estimated policy effects of large system changes, such as the introduction of the prosecutor fine in 2009 or the changes in the accessibility to subsidized legal aid, generally have had a downward effect on the forecast error.

So how do the forecast errors of the policy enriched forecasts compare to other (basic) time series models, which also could have been used? To analyze this, basic time series techniques have also been applied to key variables using the original data: keeping the forecast constant at the last known value, keeping the forecast constant at the level of the three years ahead forecast, keeping the forecast constant at the level of the four years ahead forecast, trend extrapolation of the average growth path of the last five observations, an AR(1) analysis on the last ten observations and an ARIMA(1,1,0) analysis on the last ten observations. The results show that for a period up to three years ahead the forecast errors of the policy enriched forecasts are smaller than those of alternative basic time series models, such as keeping everything constant or trend extrapolation. So in the short run our statistical model performs better. In the medium term the results are mixed. The PMJ-model performs more or less the same as keeping everything constant. In the long run keeping everything constant is actually the best option. On the whole pure time series model, such as ARIMA(p,d,q), perform worse than our statistical model.

Forecast errors are unavoidable. Also in future the accuracy delivered by the PMJ-model could be less than desirable. Continuous monitoring of both forecasts and actual developments is therefore needed to keep the forecast errors small and to avoid systematic under- or overestimation. When our statistical model was first developed, it was already noted that an explanatory model (in the statistical sense, not in a causal sense) such as this does not necessarily perform better than a basic time series model. So the fact that our statistical model seems to perform best in the short run is a bonus. The PMJ-model also has the advantage that it promotes consistency between the forecasts in various stages of the justice chains and that it is possible to use the model for simulations. However, despite all the scientific theories and research our knowledge of the determining factors that drive the need for capacity of organizations in the field of justice is still limited. Consequently, forecasts in the field of justice are not comparable to forecasts in the natural sciences. There are no natural laws which always hold and social actors within the field of justice can influence the forecast error with their own behavior. In fact, they may even react to the forecast and take measures so that the forecasts do not come true, turning the forecasts into a self-denying prophecy.

Improvements of the short term forecasts may be achieved by frequent monitoring and adjusting forecasts, by improving starting values, and by adding only policy effects of large system changes or expected breaks in trends. Easy recipes for improving medium and long term forecasts are not evident. Previous attempts of combining long term relationships (cointegration) with short term error correction models (ECM) proved to be too complicated. Another option could be to work with scenarios or confidence intervals or using different models for the near and the distant future, where the models for the distant future produce a possible outcome rather than a forecast. It is not unlikely that the distant future requires models of a different structure than the near future. The current model could be used for the near future since its performance for the near future is better than that of basic time series models. For the medium and long term the forecasts could be kept constant on the values of the last forecast of the short term model, because most organizations in the justice field need not look ahead more than one or two years. Only for prison capacity a longer planning period is required (because prisons need to be built or shut down). In this case, a long-term model may be useful.

In the near future, the advantages of using big data (such as microdata either from administrative databases or from social media, dark web or internet of things) for forecasting the first stage of the criminal justice chain seems to be limited for various reasons. Potential criminals usually do not announce their future crimes on social media. On the contrary, a smart criminal will make sure that his preparations and actions cannot be traced. Many crimes, especially violent crimes, are impulse crimes and often happen under the influence of drugs or alcohol. Often big data are not comprehensive enough, i.e. they do not cover the complete range of crimes but only a small subset. Separate big data analyses for different types of crimes seems rather cumbersome. And to forecast the need for capacity, forecasts for the explanatory factors (such as economic growth, population growth etc). need to be available for at least seven years ahead. It may be possible though to apply big data techniques to the data that are available. However, there is no guarantee for better forecasts. The use of big data for forecasting the first stage of the civil and administrative law chain seems more promising. These are typically cases where the plaintiff prepares his or her case beforehand, for example by performing some internet searches. However, even if the forecast error at the beginning of the justice chains can be improved, this does not automatically imply that forecast errors at the end of the justice chains (e.g. number of prison cells needed) can also be improved. In the end, policy choices have a substantial influence on the final outcome.