

Summary

A method to explain and predict trends

For more than a decade, annual forecasts have been made of the capacity need of the different organisations in the justice system, for the purpose of substantiating the justice budget. Since 2005, these forecasts have been made using the *Prognosemodel Justitiële ketens (PMJ, Forecasting Model for the Justice System)*. Evaluations have demonstrated, however, that the predictive quality of the PMJ, in particular in the long term, has room for improvement. One of the problems of this model is that it cannot be used to predict structural breaks in trends. This model is also not capable of dealing with breaks in trends that have occurred recently. These drawbacks indicate the need for a method that can be used to analyse trends systematically and to adequately anticipate any breaks in these trends. This report provides a description of the first steps towards a method that can do both. This method is part of a broader instrument, *Trendwatch*, which has been developed to improve the forecasts of the capacity needed in the justice system.

The method that has been developed combines two different reasoning tasks: explanation and prediction. On the one hand it will be possible to explain structural breaks in trends. On the other hand it will be possible to predict the future course of the trend based on such an explanation. This means that it will be possible to determine whether the analysed trend will continue in the same direction or whether it will change direction (in the latter case, a new break occurs). In these two forms of reasoning there are often cause and effect relationships between the entities (i.e. the entities are *causally* related). Therefore, in the *Trendwatch* method, trends and breaks in trends will be explained in terms of causal sequences of relevant *factors* (that may be changing) that shape a trend. A factor is defined as a circumstance that may occur both inside and outside the justice system.

The primary subject of the investigation is a *benchmark trend*: a current trend in the workload of an organisation in the justice system. The period in which a benchmark trend occurs is the *benchmark period*. A benchmark trend usually starts at a clearly identifiable end of a *historical trend*, that is, after a structural break in a trend. In this context it is assumed that a trend is caused by a combination of causal factors that themselves also describe trends. Factors are consequently not static, and continue to develop in the period that is being analysed. A benchmark trend is then caused by factors of which the trend has changed during the benchmark period. It is the result of a structural change in the combination of factors that shaped the historical trend.

By means of the developed *Trendwatch* method, it is possible to generate explanations for a benchmark trend. These explanations are visualised using argument diagrams. These diagrams represent the structure of an explanation in the form of a tree that consists of boxes and arrows. The boxes in this diagram represent the factors, while the arrows represent the causal relations between the factors.

The further the causal sequences are developed, the more detailed the factors included in the diagram will be. At some point in this process we do not want or are

not able to specify additional explanatory factors for these factors any further. A factor that is not explained explicitly by one or more other factors, and which is at the beginning of the causal sequence, is referred to as a *ground factor*. Still, it is important that a statement can be made about the validity of the ground factors. In the developed method, this is done by means of *arguments* based on expert opinions. For this purpose, various experts are asked whether they agree with the ground factors in a certain benchmark period. In this way, two forms of reasoning about factors are combined: causal (explanatory) and argumentative reasoning.

Firstly, causal reasoning takes place on the basis of assumed causal or explanatory relations between factors, which are established based on domain knowledge. For example:

Factor X causes factor Y or Factor X explains factor Y.

This form of reasoning makes it possible to provide a structured overview of the causes (or possible causes) of, or explanations for, a benchmark trend (and as a result of this often also for a break in a trend). In order to obtain insight into the extent to which a specific factor or ground factor explains another factor, each causal relation is assigned a value on a scale from 0 to 1. This value represents the *explanation strength* of the explaining factor in relation to the factor that has been explained. For example:

Factor X may be explained by factor Y for 0.70 (70%).

Although all factors, with the exception of the ground factors, must be explained by other factors for 100%, it is usually not possible to specify all explanations. This is why an independent factor that represents the *unknown factors* is added to each factor that has not yet been fully explained. The explanation strength of this 'unknown factor' is valued such that all explanation strengths that contribute to the factor explained by it add up to 1.

Argumentative reasoning, in the second place, is done based on expert opinions, statistics, and documents. For example:

Expert E argues that factor X is valid, so factor X is valid.

Such arguments serve to support ground factors. In this way, it is possible to give an indication of whether such a factor is valid at all. The above-mentioned argument may consequently also be read as follows:

The fact that expert E argues that factor X is valid is an indication of the validity of factor X.

Factors may be refuted in the same way, for example:

Expert E argues that factor X is not valid, so factor X is not valid.

This makes it possible to indicate that – according to a specific expert – a factor is not valid; this means that this expert is contradicting the factor. This is called a *counterargument*.

In this method, a distinction is made between factors and arguments (based on expert opinions). This difference lies primarily in the direction in which the information is represented. Relations between factors are represented in the direction from cause to effect ('factor X explains factor Y' or 'factor X causes factor Y'). Arguments based on expert opinions to support factors are represented in the opposite direction, that is, from effect to cause ('the argument of expert *E* that X is valid is an indication for factor X').

In order to be able to determine whether the factors in an explanation – and as a result of this, the explanation itself – are valid, a quantitative component has been added to the *Trendwatch* method. On the basis of the number of arguments in favour of and against a ground factor, a value is assigned to each *ground factor*. This value represents to what extent a ground factor is supported or refuted. Each argument against a ground factor is assigned the value of 0 and each argument in favour of it is assigned the value of 1. The value of a ground factor is the average of the values of all these arguments. If a ground factor has not, or not yet, been supported or refuted (consider in particular the unknown factors), it has a *default value* of 1. The idea behind this is that it is assumed that a ground factor is valid as long as there are no explicit reasons to assume that this assumption is incorrect.

If many arguments have been presented against a specific ground factor, this may be a reason to remove this ground factor from the explanation. In that case, the experts contradict that this factor is present. So, this factor may no longer be used as an explanation for subsequent factors.

When values are assigned to all ground factors and when explanation strengths of are known for all relations, it is possible to determine values for all other factors, including the benchmark trend. This is done by multiplying the sum of the values of all explanations of a factor by the explanation strength of the relations. The value that is calculated in this way represents the extent to which the factor is explained by other factors and ground factors, and thereby gives an indication of the extent to which it is present. This value is always between 0 and 1. A value of 0 implies that a factor is not present at all; a value of 1 means that the factor is indeed present. Once all these values have been calculated, it becomes clear what the validity of the explanation as a whole is. An explanation of which the value of the benchmark trend is equal to 1 has not been contradicted by experts, and is considered valid.

The thus constructed explanation for a benchmark trend is also relevant with respect to forecasting, for it can be used to predict the course of this trend in the future. In the method developed, a prediction is given of the direction (or the change of direction) of a benchmark trend in a specific *expectation period* based on the expected change of factors. An expectation period starts when the last value in the time series of the benchmark trend has been realised.

Future developments in the factors explaining a benchmark trend will determine to a significant extent what the future course of this trend will be. In the next few years, the course of the trend may:

- *Continue*: the benchmark trend continues in the same direction. This will certainly occur if all relevant factors continue to develop in the same direction as in the benchmark period.
- *Escalate*: the benchmark trend is strengthened. This is certainly the case if the trends of all relevant factors manifest themselves even more clearly.

- *Stabilise*: the benchmark trend levels out, and stabilises at the current level. This will certainly occur if the trends of all relevant factors stabilise at the current level.
- *Reverse*: the benchmark trend develops in the opposite. This is certainly the case if all trends contained in the relevant factors also reverse.

In order to be able to determine in which direction a benchmark trend will develop, it must also be determined in which direction all factors in the established explanation will develop. Expert opinions will once again form the basis of this process. Experts will be asked to assess how they think the ground factors will develop in the future. These expert opinions will subsequently be expressed in arguments with a *prediction direction*. For this purpose, there are four scenarios, each with its own value: a factor may continue (prediction direction=1), reverse (-1), stabilise (0), or escalate (2).

Based on these arguments a *prediction value* for each *ground factor* can be calculated. These values are between -1 and 2. The prediction value of a ground factor is the average of the prediction directions of all its arguments, where ground factors without arguments are assigned a default value of 1. Once the prediction values for all ground factors has been determined, it is also possible to assign prediction values to the other factors in the argument diagram. This is done by multiplying the sum of the prediction values of all explanations of a factor by the explanation strength of the relations.

On the basis of this calculation, it will subsequently be possible to calculate a prediction value for the benchmark trend. This prediction value gives a *trend indication*: this is a value that indicates whether and, if so, in which direction the benchmark trend will change in a specific expectation period. On the basis of this trend indication, it is possible to determine the expected *trend level* (of the workload of an organisation) at the end of the expectation period.

A trend indication of 1 points to a continuation of the benchmark trend. In that case, a new structural break is not expected in the expectation period. A value of 0, however, has been defined as a stabilisation of the benchmark trend around the level of the last realisation. This means a break in each situation in which the benchmark trend itself is not stable.

As the trend level at the end of an expectation period is known for both a trend indication of 0 (this is the same level as in the last year of the benchmark period) and 1 (this can be calculated by extrapolation of the benchmark trend), the expected trend level can be calculated reasonably accurately for all trend indications between 0 and 1. In this way, it will be possible to come to a well-founded prediction based on expert opinions. This prediction may subsequently be used to test and, where required, adjust forecasts of the capacity needed in the justice system.