# Some Remarks on the Method of Cross-Influence of Events

Sławomir Dorosiewicz
*Motor Transport Institute and Warsaw School of Economics*

The sufficient condition of existence and uniqueness of forecasts made by the method of cross-influence of events is formulated and proved. This result is ilustrated by the forecasts of the volume of road transport in Poland up to the year 2025.
**Keywords:** Transport, Forecasting, Cross-influence of events.

## 1. CONSTRUCTION OF THE FORECASTS IN THE METHOD OF CROSS-INFLUENCE OF EVENTS

The method of cross influence of events (CIE, [4], [10]) is, except such methods as Brain Storm or Delphi Method, one of widely known and commonly used forecasting procedures. The high popularity of this method derives from its flexibility which allows it to use expert's knowledge, experience and intuition, and also their ability to join ordering various kinds of information and data. These features obviously desirable and indispensable take on special importance in all situations where no quantitative data exist or is incomplete. The CIE method is especially popular in the construction of long-term forecasts when the possibility of occurrence of structural changes within the forecast time-horizon makes many other methods unsuitable[1]. It seems that the CIE method should be regarded as an alternative to other forecasting procedures, not standing in opposition to them, but rather as a complementary procedure. This paper provides some remarks on the problem of existence and uniqueness of forecast made by CIE procedure.

Application of CIE method requires definitions of the following main categories:

scenarios and their probabilities;
forecasted cathegory (quantity, variable) and its values for different scenarios.

If these elements are specified correctly, then the forecasted cathegory is a random variable. Its probability distribution defines the pointwise and interval forecasts and all their characteristics.

Let $Z$ denote a set of factors which are important in forecasting process, ie. the factors which potentially influence forecasted variable $x$. For simplicity we shall assume that $Z$ is a finite set. Any subset of $Z$ will be called *a scenario,* a family $S = 2^Z$ of all subsets of $Z$ defines the family of all possible scenarios.

Clearly the value of forecasted variable ($x$) may differ across scenarios. Formally this variable is a map $x: S \to R$. One of the main difficulties of applying CIE method arise from large number of factors potentially impacting the forecasting variable. It is usually difficult to require an expert to define the values of function $x$ for all scenarios. If it is possible to extract the individual impact of particular factors on the forecast,

[1]It is difficult to present even main applications of CIE method in various kind of forecasting. Some examples may be found in [2,3,6,9,11].

definition of $x$ can be written in the following additive form:

$$\forall\, s \in S \;:\qquad x(s) = \sum_{z \in s} x_z + \sum_{z \notin s} x_{-z}, \qquad (1)$$

where $x_z$ (resp. $x_{-z}$) measures the impact the presence (resp. absence) of factor $z$ (ceteris paribus) on forecasted variable[2]. Acceptance of this assumption reduces the experts' effort: they have to define only values of $x_z$ and $x_{-z}$, that is only $2|Z|$ values (usually $2|Z| \ll |S| = 2^{|Z|}$).

To determine the forecast of $x$ the probability measure defined on whole family of scenarios is needed. Unfortunately the experts' defined probabilities are not usually consistent, so they do not satisfy the formal requirements of probability theory. The second main difficulty of applying CIE method is to find a measure which is the probability in a strictly formal sense and which is also possibly closest to experts' probabilities. It is easy to check that any nonnegative map $\pi : S \to R$ satisfying condition $\sum_{s \in S} \pi(s) = 1$, defines a probability over the family of all scenarios:

$$\Pi_\pi : 2^S \to R, \quad \Pi_\pi(A) = \sum_{s \in A} \pi(s) \qquad (2)$$

We discuss the problem how to find $\pi$ which "best matches"' to expert's data.

Let $\mathbf{E}$ be a set of all pairs $(A, B) \in 2^S \times 2^S$ for which experts are able to specify conditional probabilities $p(A \mid B)$ of $A$ if $B$ holds. Each distribution (2) allows to compute these probabilities. Namely, the value of probability of $A$ if $B$ occurs is equal:

$$\Pi_\pi(A \mid B) = \Pi_\pi(A \cap B) / \Pi_\pi(B) = \sum_{s \in A \cap B} \pi(s) \Big/ \sum_{s \in B} \pi(s). \qquad (3)$$

The main problem is to choose $\Pi_\pi$ properly, to make probabilities (3) possibly closest to $p(A \mid B)$ for all pairs from $\mathbf{E}$. A quite natural criterion of goodness of fit may have the following form:

$$F_1(\pi) := \sum_{(A,B) \in \mathbf{E}} \left( p(A \mid B)\Pi_\pi(B) - \Pi_\pi(A \cap B) \right)^2 \qquad (4)$$

or

$$F_2(\pi) := \sum_{(A,B) \in \mathbf{E}} \left( p(A \mid B) - \Pi_\pi(A \mid B) \right)^2, \qquad (5)$$

subject to the following constraints:

$$\sum_{s \in S} \pi(s) = 1, \qquad \forall\, s \in S \;:\; \pi(s) \geq 0.$$

Those constraints define a unit $|S|$-dimensional simplex, which will be denoted by $\Delta$

Both functions (4) and (5) measure deviation between probabilities evaluated by experts and probabilities computed using a distribution $\Pi$. It is easy to see that first of those functions can be rewritten as:

$$F_1(\pi) := \sum_{(A,B) \in \mathbf{E}} \Pi_\pi^2(B) \left( p(A \mid B) - \Pi_\pi(A \mid B) \right)^2. \qquad (7)$$

This means that $F_1$ refers to the sum of weighted deviations of conditional probabilities. The weights are obviously proportional to the squares of absolute probabilities ($\Pi_\pi(B)$). The second function is the unweighted sum of conditional probabilities, so it threats all deviations in the same way independently how probable are consecutive events. This is the main reason that the function $F_2$ seems to define "a worse" criterion of choosing probability $\Pi_\pi$ than the criterion defined by $F_1$. In the next part of this paper we will consider function (5). Fortunately this map has some very desired properties. One of them is a convexity.

**Theorem.** The map (4) is a convex function of $\pi$. This map is strictly convex if

$$\ker \phi \cap H = \{0\},$$

where $\phi$ is the (linear) map $\phi : R^{|S|} \to R^{|\mathbf{E}|}$

---

[2]In this formulation of CIE method experts should define the "ordinary" values of $x$. In some reformulations of this method fuzzy numbers are allowed ([1]).

$$x \mapsto \left( p(A\,|\,B)\sum_{s\in B} x(s) - \sum_{s\in A\cap B} x(s): \quad (A,B)\in \mathbf{E} \right) \tag{8}$$

and $H$ is the hiperplane given by $\sum_{s\in S} x_s = 0$.

**Proof.** First observe that continuity of $F_1$ and compactness of the simplex $\Delta$ follow that the problem of minimizing $F_1(\pi)$ subject to $\pi \in \Delta$ has at least one solution. The map (4) can be written as $F_1(\pi) = f(\phi(\pi))$, where $f(x) = \sum_{e\in\mathbf{E}} x_e^2$ and $\phi$ is defined by (8). This means that $F_1$ is a superposition of strictly convex map and a linear one, so $F_1$ is convex. It remains to prove strict convexity of $F_1$. Let $t \in [0,1], \pi_1, \pi_2 \in \Delta$. The equity

$$F_1(t\pi_1 + (1-t)\pi_2) = tF_1(\pi_1) + (1-t)F_1(\pi_2)$$

holds if and only if $t = 0$ or $t = 1$ or $\phi(\pi_1) - \phi(\pi_2) = 0$. It follows from linearity of $\phi$ that in the last case $\pi_1 - \pi_2 \in \ker\phi$. Because of $\phi_1, \phi_2 \in \Delta$ we have $\phi_1 - \phi_2 \in H$ and consequently $\phi_1 - \phi_2 \in \ker\phi \cap H$. Finally $\phi_1 = \phi_2$, so the map $F_1$ is strictly convex. The proof is complete.

Strictly convex function has only one minimum, so under assumption of Theorem 1, the problem of maximizing function (4) subject to $\pi \in \Delta$ has exactly one solution. There exists only one random distribution which is best fitted to experts data. In order to have this, the number $|\mathbf{E}|$ should be large, i.e. the experts should be able to estimate sufficiently large number of probabilities $p(A\,|\,B)$. This unique measure defines various kinds of forecasts, for example mean-square and quantile ones.

A **pointwise mean-square forecast** $\mu$ has the property

$$\forall\, c \in R \quad E_\pi(x-\mu)^2 \le E_\pi(x-c)^2. \tag{9}$$

It can be easily shown that this forecast is an expected value of $x$:

$$\mu = E_\pi x = \sum_{s\in S} \pi(s)x(s). \tag{10}$$

For **quantile forecast** $m$ we have:

$$\forall\, c \in R \quad E_\pi\,|x-m| \le E_\pi\,|x-c|; \tag{11}$$

it is not very hard to show that $\hat{x}^{(2)}$ is equal the median of $x$, i.e. any number $m$ satisfying the condition

$$\Pi_\pi(x < m) \le 1/2 \le \Pi_\pi(x \le m). \tag{12}$$

A symmetric interval forecast for significance level $\alpha \in [0,1]$ is an interval $[x^-, x^+]$, where $x^-$, $x^+$ are respectively upper $(1-\alpha)/2$ - and lower $(1+\alpha)/2$ quantile of $\Pi_\pi$. Consequently this interval forecast satisfies the inequality

$$\Pi_\pi(x \in [x^-, x^+]) \ge \alpha.$$

In one of the most important case of this method experts estimate the conditional probabilities for **all** pairs of factors, i.e. all the probabilities $p(S_x\,|\,S_y)$, $p(S_x\,|\,S_{-y})$, $p(S_x\,|\,S)$, where $S_z$ denotes the family of all scenarios containing an event $z$, $S_{-z}$ is the complement of $S_z$ to $S$, that is $S_{-z} = S - S_z$. If the number of probabilities estimated by experts is large enough, then theorem 1 holds. This kind of situation takes place in the exemplary forecast from the next section.

## 2. A SPECIAL CASE: THE FORECAST OF VOLUME OF ROAD TRANSPORT IN POLAND

In this section we apply the method of the cross influence of events to compute forecast of the volume of road transport in Poland in forthcoming 15 years, i.e. till 2015. The results are supplementary to the earlier forecasts (see for example [8],[4]).

The first stage of the construction is the process of selecting of factors which play the most important role in the development of road transport in Poland[3]. The most important factors which have been chosen for making forecast are the following:

$a$: growth of polish economy (measured by GDP),

$b$: growth of european and world economy,

[3] I would like to thank my colleagues from Motor Transport Institute for making the data available.

- $c$: growth of consumption,
- $d$: increasing efficiency in transport resulting from development of informatics and communication technologies,
- $e$: influence the government politics on transport,
- $f$: development and modernization of transport infrastructure,
- $g$: changes of other factors significant for transport sector.

We have $Z=\{a,b,\ldots,g\}$, $S=2^Z$, hence there are $|S|=2^7$ possible scenarios in this model. Experts were requested to estimate the following data: 1. probabilities $p(S_i\mid S_j), p(S_i\mid S_{-j}), p(S_i\mid S)$ for all pairs $(i,j)\in Z\times Z$ and 2. the influence of every factor $i\in Z$ on the volume of transport (measured in tonne-kilometers) in consecutive years up to 2025. The mentioned influence is measured by the annual rates of change of transport volume. The basis year is 2009 with volume of transport equal 100%. A forecasted variable is a volume (measured in tonne-kilometers, year 2009 = 100%) of road transport in Poland. Its value in year $t$ and for a given scenario $s$ is equal:

$$x_t(s)=\prod_{\tau=2010}^{t}\prod_{s\in S}(1+r_z(t)m_z(s)),\quad \forall t=2010,\ldots,2025,$$

where $m_z(s)=1$ if an event $z$ occurs in scenario $s$ (i.e. $z\in s$), otherwise $m_z(s)=0$, $r_z(t)$ is the rate of change volume of transport in the $t$ as a result of the factor $z$ (see table A3 in Appendix).

A pointwise mean-square forecasts, equal the expected value of $x_t$,

$$\mu_t=E_\pi x_t=\sum_{s\in S}x_t(s)\pi(s),$$

are shown in table 1. The column labeled "$\sigma_t$" includes values of standard deviations of forecasts:

$$\sigma_t=\sqrt{E_\pi(x_t-\mu_t)^2}=\sqrt{\sum_{s\in S}(x_t(s)-\mu_t)^2\pi(s)}.$$

Another characteristic is forecast band. Its lower and upper limit for time $t$ is equal $\mu_t-2\sigma_t$ and $\mu_t+2\sigma_t$ respectively. The last column contains probabilities that value of $x_t$ belongs to $[\mu_t-2\sigma_t,\mu_t+2\sigma_t]$.

The year 2009 with the transport volume 100% is the basis. In the next year the forecasted volume of transport with probability 0.965 should be between from 99.86% to 104.58%. Although the volume of transport may descend (in comparison to basis year, see the low band below 100%), but it is very likely that we shall observe increasing volume and an upward trend in next couple of years. In the year 2025 the volume of transport with a large probability should belong to the interval [93.4%, 204.6%], i.e. its mean is almost 50% greater than in the basis year.

Table 1: Mean-square forecast and its characteristics, year 2009 = 100

| Year | $\mu_t$ | $\sigma_t$ | $\mu_t-2\sigma_t$ | $\mu_t+2\sigma_t$ | probability |
|---|---|---|---|---|---|
| 2010 | 102.22 | 1.18 | 99.86 | 104.58 | 0.965 |
| 2011 | 104.51 | 2.41 | 99.69 | 109.33 | 0.965 |
| 2012 | 106.86 | 3.70 | 99.47 | 114.25 | 0.963 |
| 2013 | 109.28 | 5.03 | 99.21 | 119.35 | 0.963 |
| 2014 | 111.77 | 6.43 | 98.91 | 124.64 | 0.959 |
| 2015 | 114.33 | 7.89 | 98.55 | 130.11 | 0.959 |
| 2016 | 117.46 | 9.50 | 98.45 | 136.46 | 0.974 |
| 2017 | 120.68 | 11.20 | 98.29 | 143.08 | 0.974 |
| 2018 | 124.02 | 12.99 | 98.04 | 150.00 | 0.974 |
| 2019 | 127.47 | 14.87 | 97.72 | 157.21 | 0.974 |
| 2020 | 131.03 | 16.85 | 97.32 | 164.74 | 0.974 |
| 2021 | 134.40 | 18.84 | 96.73 | 172.08 | 0.981 |
| 2022 | 137.89 | 20.92 | 96.04 | 179.73 | 0.981 |
| 2023 | 141.48 | 23.11 | 95.26 | 187.70 | 0.981 |
| 2024 | 145.19 | 25.41 | 94.37 | 196.00 | 0.981 |
| 2025 | 149.01 | 27.81 | 93.38 | 204.64 | 0.981 |

Source: own calculations.

Table 2 contains computed quantile forecast and some of its characteristics. The columns show 0%, 25%, 50%(median), 75% and 100% quantiles of forecast. Median column defines the pointwise quantile forecast, minimum and maximum columns define minimal ($\min_{s\in S}x_t(s)$) and maximal ($\max_{s\in S}x_t(s)$) forecast respectively.

In 2011 the minimal volume of transport should be 99.43% in comparison to basis year. With probabilities $1/4$ and $3/4$ transport volume changes no more than 2.92% and 6.28% respectively. The maximum rate of change should not increase 110.23%. In the last year with probabilities $1/4$ and $3/4$ forecasted variable should not excess 129% and 167.77%. Like in mean-square forecast the traffic volume at the end of the period should not exceed 220-230% of its base value.

Table 2: Quantile forecast and its characteristics (year 2009=100)

| Year | minimum | 25% quartile | median | 75% quartile | maximum |
|---|---|---|---|---|---|
| 2010 | 99.43 | 101.45 | 102.27 | 103.09 | 104.99 |
| 2011 | 98.86 | 102.92 | 104.59 | 106.28 | 110.23 |
| 2012 | 98.30 | 104.41 | 106.97 | 109.56 | 115.73 |
| 2013 | 97.74 | 105.93 | 109.39 | 112.94 | 121.50 |
| 2014 | 97.18 | 107.46 | 111.88 | 116.43 | 127.57 |
| 2015 | 96.63 | 109.02 | 114.42 | 120.03 | 133.93 |
| 2016 | 96.19 | 110.87 | 117.05 | 124.26 | 141.59 |
| 2017 | 95.76 | 112.76 | 119.74 | 128.63 | 149.69 |
| 2018 | 95.33 | 114.68 | 122.92 | 133.16 | 158.26 |
| 2019 | 94.90 | 116.69 | 126.27 | 137.85 | 167.31 |
| 2020 | 94.47 | 118.75 | 129.70 | 142.70 | 176.88 |
| 2021 | 93.86 | 121.13 | 132.65 | 147.02 | 186.50 |
| 2022 | 93.25 | 123.13 | 135.67 | 151.59 | 196.65 |
| 2023 | 92.64 | 124.50 | 138.77 | 156.22 | 207.34 |
| 2024 | 92.04 | 126.74 | 141.93 | 161.77 | 218.62 |
| 2025 | 91.44 | 129.02 | 145.17 | 167.77 | 230.52 |

Source: own calculations.

## 3. APPENDIX. THE DATA

Tables A1-A3 include necessary data (from Motor Transport Institute, Department of Economic Research) for computing forecasts.

Table A1. Conditional probabilities (rows $a-g$). First row contains probabilities $p(a\,|\,a)=1,\ldots,p(a\,|\,g)$. Next six rows have analogous meanings. Last row consists of probabilities $p(a),\ldots,p(g)$. Each value is an average of values given by 12 experts.

| Events | a | b | c | d | e | f | g |
|---|---|---|---|---|---|---|---|
| a | 1.00 | 0.46 | 0.57 | 0.41 | 0.43 | 0.49 | 0.42 |
| b | 0.55 | 1.00 | 0.53 | 0.47 | 0.31 | 0.49 | 0.43 |
| c | 0.53 | 0.51 | 1.00 | 0.39 | 0.29 | 0.45 | 0.37 |
| d | 0.52 | 0.45 | 0.40 | 1.00 | 0.46 | 0.45 | 0.39 |
| e | 0.33 | 0.26 | 0.32 | 0.36 | 1.00 | 0.40 | 0.39 |
| f | 0.59 | 0.49 | 0.47 | 0.42 | 0.36 | 1.00 | 0.36 |
| g | 0.36 | 0.38 | 0.35 | 0.35 | 0.39 | 0.37 | 1.00 |
| Probabili-ty | 0.48 | 0.49 | 0.49 | 0.56 | 0.49 | 0.57 | 0.46 |

Table A2. Conditional probabilities continued. Averaged expert's evaluations. First row contains probabilities

$$p(a\,|\,\mathbf{no}\;a)=0,\ldots,p(a\,|\,\mathbf{no}\;g)$$

($\mathbf{no}\;z$ means that an event $z$ does not occur). Next rows have analogous meanings). Average of expert's evaluations.

| Events | no a | no b | no c | no d | no e | no f | no g |
|---|---|---|---|---|---|---|---|
| a | 0.00 | 0.15 | 0.14 | 0.43 | 0.36 | 0.33 | 0.34 |
| b | 0.48 | 0.00 | 0.41 | 0.50 | 0.56 | 0.46 | 0.48 |
| c | 0.11 | 0.21 | 0.00 | 0.45 | 0.53 | 0.47 | 0.49 |
| d | 0.29 | 0.39 | 0.38 | 0.00 | 0.40 | 0.34 | 0.28 |
| e | 0.37 | 0.40 | 0.41 | 0.36 | 0.00 | 0.31 | 0.45 |
| f | 0.18 | 0.23 | 0.26 | 0.28 | 0.32 | 0.00 | 0.36 |
| g | 0.31 | 0.33 | 0.29 | 0.32 | 0.32 | 0.35 | 0.00 |

Table A3. Rates of change of goods road transport (in t-km) as the results of occurence of particular event (i.e. the values of $x_z$ in formula (1); values of $x_{-z}$ are assumed to be 0)). Average of expert's evaluations.

| Event | Period | | |
|---|---|---|---|
| | 2010-2015 | 2016-2020 | 2021-2025 |
| a | 1.24 | 1.23 | 1.36 |
| b | 1.21 | 1.14 | 1.15 |
| c | 1.03 | 1.07 | 0.58 |
| d | 0.91 | 1.12 | 1.05 |
| e | -0.57 | -0.45 | -0.65 |
| f | 0.18 | 0.53 | 0.59 |
| g | 0.42 | 0.63 | 0.71 |

## LITERATURE

[1] Asan U., Bozdag C.E., Polat S., A fuzzy approach to qualitative cross-impact analysis, Omega 32, 2004.

[2] R.W. Blanning, B.A. Reinig, Cross-impact analysis using group decision support systems: an application to the future of Hong Kong, Futures 31 (1), 1999.

[3] Choi C., Kim S., Park Y., A patent-based cross impact analysis for quantitative estimation of technological impact: The case of information and communication technology, Technol. Forecast. Soc. Change, vol. 74, 2007.

[4] Cieślak M. Prognozowanie gospodarcze. Metody i zastosowania. PWN, Warszawa 1997r.
[5] Bereziński M. O jednej z metod badania wpływu narzędzi ekonomicznych na rynek usług transportowych. Opracowanie w ramach projektu badawczego KBN, nr 909452-92-03. ITS, Warszawa 1994.
[6] Dorosiewicz S., Dorosiewicz T. Some remarks on the method of cross influence of events. Application to forecasting of transport developments. Archives of Transport, vol 12, Issue 1, 2000r.
[7] Enzer S, Delphi and cross-impact techniques: An effective combination for systematic futures analysis, Futures 3, 1971.
[8] Gordon T.J., Hayward H. Initial Experiments with the Cross-Impact Matrix Method of Forecasting. Futures, vol. 1, no 2, 1962.
[9] Jeong G.H., Kim S.H., A qualitative cross-impact approach to find the key technology, Technol. Forecast. Soc. Change 55, 1997.
[10] Makridakis S., Wheelwright S.C., Forecasting Methods for Management, New York, J. Wiley and Sons, 1989.
[11] Sveshnikov S., Bocharnikov V., Forecasting Financial Indexes with Model of Composite Events Influence, Journal of Applied Economic Sciences, Volume IV, Issue 3(9), 2009.

**Sławomir Dorosiewicz**
**Szkoła Główna Handlowa**
**doro@sgh.waw.pl**