

Fully Flexible Bayesian Networks¹

Attilio Meucci²

attilio_meucci@symmys.com

this version: December 7, 2010

latest version available at <http://ssrn.com/abstract=1721302>

Abstract

We propose a methodology to stress-test a set of risk drivers under minimal information. This methodology applies the entropy-based "fully flexible views" approach in Meucci (2008) and a novel consistency algorithm to extend the Bayesian network approach in Rebonato (2010). Starting from a plausible market distribution of the drivers, we are able to stress any conditional probabilities. These conditional probabilities comprise, but are not restricted to, causal Bayesian networks. Furthermore, in our approach we can also stress-test expectations, volatilities, correlations, quantiles, medians, etc.

We detail the theory and we present a case study: stress-testing a market driven by swap curve, credit spreads, stock market return, stock market volatility, currency strength, inflation, and commodities. Fully commented code supporting the empirical analysis is available for download.

JEL Classification: C1, G11

Keywords: Bayesian network, entropy pooling, non-Boolean variables, prior distribution, posterior distribution, linear programming, convex programming, dual optimization.

¹Preliminary version under submission. Feedback of any nature is very welcome

²The author is grateful to Davide Di Gennaro and to an anonymous referee

1 Introduction

The distribution of the key risk drivers in any given market can never be estimated correctly. Therefore, stress-testing becomes the only effective tool to handle estimation risk both in risk management and in portfolio management: the base-case risk model is modified manually and the ensuing p&l distribution is computed and evaluated.

In turn, stress-testing presents two difficulties. First, it is hard to specify reasonable modifications of a base-case risk model. Second, it is challenging to embed such modifications coherently in the base-case risk model.

The entropy-based "fully flexible views" approach in Meucci (2008) solves the latter issue in full generality: under arbitrary distributional assumption for the base-case risk model, the stressed model is defined as the one that displays the least distortion from the base-case risk model and yet satisfies the stress-tests, where the distortion is measured by the relative entropy.

To cope with the first problem, namely the difficulty to specify reasonable stress-tests, Rebonato (2010) proposes to greatly simplify the structure of the risk model in two directions. First, the risk drivers are modeled as Boolean variables, which can only take on the values "true" or "false". Second, such risk drivers are assumed to be connected by a parsimonious causal Bayesian network: this network defines the risk drivers distribution in terms of a limited number of conditional probabilities. This double layer of simplification, namely true/false events and sparse causal network, reduces stress-testing to re-assessing a manageable number of conditional probabilities, of which the practitioner is supposed to have a deep understanding.

Here we use fully flexible views to extend Rebonato (2010). More precisely, any situation covered by Rebonato (2010) is also covered by the current approach as a special case, but extensions are provided in four directions.

First, we do not impose that the risk drivers only assume "true" or "false" values. Instead, we let them take on more realistic arbitrary sets of discrete outcomes.

Second, we do not need to assume that the risk model is a Bayesian network. Instead, we stress-test arbitrarily complex causal conditional probabilities, and we recover any unstressed feature by minimum-entropy arguments.

Third, we allow for additional direct stress-testing of correlations, volatilities, expectations, quantiles, and other features of the risk drivers.

Finally, we provide a fully general algorithm to ensure that the stress-tests are consistent with each other. This algorithm, unlike the one in Rebonato (2010) applies to all possible assessments of the probabilities of the scenarios, conditioned on any number of statements.

In Section 2 we present our theoretical framework. We review entropy pooling, highlighting how it covers generalized stress-testing of conditional probabilities under discrete scenarios. Then we introduce our consistency algorithm for the stress-tests. Throughout the discussion we illustrate the theory by means of a toy example.

In Section 3 we present a real-life case study. We consider a market driven

by swap curve, credit spreads, stock market return, stock market volatility, currency strength, inflation, and commodities. We stress-test this market by assessing a few key conditional and marginal probabilities as well as correlations. Then we compare the prior and the posterior distribution that follow from those stress-tests.

Fully commented code supporting the empirical analysis is available for download.

2 Entropy pooling on conditional distributions

This section is adapted from the general fully flexible approach in Meucci (2008): first, we define the prior, base-case distribution for the risk drivers in terms of a panel of joint scenarios \mathcal{X} and their respective probabilities \mathbf{p} . Then we formulate the stress-tests of the base-case in terms of linear constraints on a yet-to-be defined posterior set of probabilities $\tilde{\mathbf{p}}$. Next, we check for the consistency of such constraints. Finally, we define the stressed posterior distribution as the probabilities $\tilde{\mathbf{p}}$ that display the least relative entropy from the prior and at the same time satisfy the stress-test constraints.

The prior

We consider N risk drivers $\mathbf{X} \equiv (X_1, \dots, X_N)'$, namely random variables that determine the p&l of the portfolio under scrutiny. We assume that such drivers are discrete, i.e. for each $n = 1, \dots, N$ the driver X_n can only result in a finite set of K_n outcomes.

$$X_n \in \{x_{n,1}, \dots, x_{n,K_n}\}. \quad (1)$$

For instance, consider the case of $N \equiv 3$ drivers X_1, X_2, X_3 . Say X_1 describes the widening of spreads, which can take values "low", "medium", and "high"; X_2 describes the default of a given country, which can take on values "default" or "survival"; and X_3 describes the central bank intervention, which can either "cut" or "raise" interest rates. Therefore in the formalism (1) we obtain

$$X_1 \in \{L, M, H\} \quad (2)$$

$$X_2 \in \{D, S\} \quad (3)$$

$$X_3 \in \{C, R\}. \quad (4)$$

The total number of joint scenarios is $J \equiv \prod_{n=1}^N K_n$. Rebonato (2010) discusses a special case of this framework, where each driver is Boolean, i.e. $X_n \in \{T, F\}$ and thus there are $J \equiv 2^N$ possible joint scenarios.

We collect all the J joint scenarios of the N risk drivers \mathbf{X} in a $J \times N$ panel \mathcal{X} . The stochastic properties of these events are fully described by the J -dimensional vector \mathbf{p} of the probabilities of each joint scenario. We call $(\mathcal{X}, \mathbf{p})$ the "prior" distribution of the market.

In our example, there are $J = 3 \times 2 \times 2 = 12$ joint scenarios. Assume that all the joint events are equally probable. Then the prior becomes

\mathcal{X}			\mathbf{p}
L	D	C	$1/12$
L	D	R	$1/12$
L	S	C	$1/12$
L	S	R	$1/12$
M	D	C	$1/12$
M	D	R	$1/12$
M	S	C	$1/12$
M	S	R	$1/12$
H	D	C	$1/12$
H	D	R	$1/12$
H	S	C	$1/12$
H	S	R	$1/12$

(5)

Rebonato (2010) determines the prior probabilities by imposing a parsimonious Bayesian network structure. Mathematically, this amounts to replacing the full specification of the joint probability $\mathbb{P}\{x_1, \dots, x_N\}$ with a parsimonious conditional representation

$$\mathbb{P}\{x_1, \dots, x_N\} \equiv \prod_{n=1}^N \mathbb{P}\{x_n | \mathbf{x}_{c(n)}\}, \quad (6)$$

where $\mathbf{x}_{c(n)}$ denotes the set of variables among (x_1, \dots, x_N) that have a causal effect on the driver X_n .

For instance assume that in our example the spread changes X_1 are caused by the central bank action X_3 and nothing else. Then the following Bayesian network is defined: $\{X_2; X_1 \leftarrow X_3\}$. Therefore the entries of the probability vector \mathbf{p} in (5) can be generated as follows

$$\mathbb{P}\{x_1, x_2, x_3\} = \mathbb{P}\{x_1 | x_3\} \mathbb{P}\{x_2\} \mathbb{P}\{x_3\}. \quad (7)$$

To generate all the $J = 12$ probabilities, only 6 numbers must be specified, namely $\mathbb{P}\{X_2 = D\}$, $\mathbb{P}\{X_3 = C\}$, $\mathbb{P}\{X_1 = L | X_3 = C\}$, $\mathbb{P}\{X_1 = M | X_3 = C\}$, $\mathbb{P}\{X_1 = L | X_3 = R\}$, and $\mathbb{P}\{X_1 = M | X_3 = R\}$.

More in general, we can assign the prior distribution using classical frequentist analysis, as in the case study below. Alternative approaches are also possible, we refer the reader to the vast literature on estimation theory.

The stress-tests

A stress-test is a subjective statement on features of the distribution of the risk drivers \mathbf{X} . Typically, risk managers stress-test expectations, correlations and

volatilities of the risk drivers. However, Rebonato (2010) proposes to stress-test the conditional probabilities that define the Bayesian network, under the rationale that practitioners have a better grasp on their market in relative-causal terms than in absolute terms.

More in general, we stress-test any generic conditional probability. Denoting by $\tilde{\mathbb{P}}$ the subjective probability of the practitioner, the generic k -th view is in the format

$$\tilde{\mathbb{P}}\{X_{n_1} \in \mathbf{x}_{n_1}, X_{n_2} \in \mathbf{x}_{n_2}, \dots | X_{m_1} \in \mathbf{x}_{m_1}, X_{m_2} \in \mathbf{x}_{m_2}, \dots\} \geq \tilde{v}_k, \quad (8)$$

where \mathbf{x}_n denotes a subset of the potential outcomes (1) of the risk driver X_n , i.e. $\mathbf{x}_n \subset \{x_{n,1}, \dots, x_{n,K_n}\}$; \geq denotes any (in)equality; and \tilde{v}_k is a subjective probability threshold.

We emphasize that our framework (8) allows for multiple outcomes in the conditional probabilities. Furthermore, we do not require the existence of a Bayesian network.

For instance, consider the following stress-test: the probability of the foreign country defaulting is at least 30%, whereas the probability of spreads widening to a moderate or large extent if the foreign country defaults is at least 70%. In formulas, this reads

$$\tilde{\mathbb{P}}\{X_1 \in \{M, H\} | X_2 = D\} \geq 0.7. \quad (9)$$

$$\tilde{\mathbb{P}}\{X_2 = D\} \geq 0.3 \quad (10)$$

Notice that in the first stress-test we are letting the variable X_1 take on multiple values, namely M and H , and we are not stress-testing the probabilities $\tilde{\mathbb{P}}\{x_1|x_3\}$, and $\tilde{\mathbb{P}}\{x_3\}$ that define the Bayesian network (7).

Each view/stress-test (8) can be rephrased in terms of linear constraints on the vector $\tilde{\mathbf{p}}$ of subjective probabilities which are associated with each scenario. Indeed, denote by $\mathbf{I}_{jnt(k)}$ the indicator of the scenarios where the joint conditions in the stress-test (8) are satisfied and denote by $\mathbf{I}_{cnd(k)}$ the indicator of the scenarios where the conditioning events in the stress-test (8) are satisfied. Then using the identity $\mathbb{P}\{A|B\} = \mathbb{P}\{A \cap B\} / \mathbb{P}\{B\}$ we can reformulate (8) as

$$(\mathbf{I}_{jnt(k)} - \tilde{v}_k \mathbf{I}_{cnd(k)})' \tilde{\mathbf{p}} \geq 0. \quad (11)$$

Also, In the special case where there is no conditioning statement, (8) becomes

$$\mathbf{I}'_{jnt(k)} \tilde{\mathbf{p}} \geq \tilde{v}_k. \quad (12)$$

Both (11) and (12) are linear constraints on $\tilde{\mathbf{p}}$.

Consider our example (9), where

$$\mathbf{I}_{jnt} \equiv \mathbf{I}_{X_1 \in \{M, H\} \cap X_2 = D} \quad (13)$$

$$\mathbf{I}_{cnd} \equiv \mathbf{I}_{X_2 = D}. \quad (14)$$

From (5) we obtain

$$\mathbf{I}_{jnt} = (0, 0, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0)' \quad (15)$$

$$\mathbf{I}_{cnd} = (1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0)' . \quad (16)$$

Therefore, according to (11), the stress-test (9) reads

$$-0.7\tilde{p}_1 - 0.7\tilde{p}_2 + 0.3\tilde{p}_5 + 0.3\tilde{p}_6 + 0.3\tilde{p}_9 + 0.3\tilde{p}_{10} \geq 0. \quad (17)$$

Similarly, for (10) we obtain

$$\tilde{p}_5 + \tilde{p}_6 + \tilde{p}_7 + \tilde{p}_8 + \tilde{p}_9 + \tilde{p}_{10} \geq 0.3. \quad (18)$$

Furthermore, as we show in Meucci (2008), stress-tests on expectations, volatilities, and correlations can also be expressed as linear constraints on the probabilities. Therefore, all the views/stress-tests can be summarized as

$$\mathbf{A}\tilde{\mathbf{p}} \leq \mathbf{b}, \quad (19)$$

where \mathbf{A} and \mathbf{b} are a suitable conformable matrix and vector respectively.

Consistency check

The views/stress-tests might not be consistent, i.e. there might not exist a vector \mathbf{p} that satisfies them and is also a probability vector

$$\mathbf{1}'\mathbf{p} \equiv 1, \quad \mathbf{p} \geq \mathbf{0}. \quad (20)$$

In Rebonato (2010) the problem is solved under specific circumstances, namely when the stress tests are performed on conditional distributions that do not involve more than a few conditioning variables. Here, we propose an algorithm that works in full generality.

In our approach, in order to guarantee the consistency of the stress-tests, we relax them

$$\mathbf{b} \mapsto \mathbf{b} + \delta\mathbf{b}, \quad (21)$$

where $\delta\mathbf{b} \geq \mathbf{0}$ is a minimal perturbation that in general is null $\delta\mathbf{b} \equiv \mathbf{0}$, but that could become strictly positive in some or all of the entries. To compute $\delta\mathbf{b}$ we solve the following problem

$$\delta\mathbf{b} \equiv \underset{\mathbf{x}}{\operatorname{argmin}} \{ \boldsymbol{\gamma}'\mathbf{x} \} \quad (22)$$

where \mathbf{x} satisfies

$$\mathbf{x} \geq \mathbf{0} \quad (23)$$

$$\mathbf{A}\mathbf{p} \leq \mathbf{b} + \mathbf{x} \quad (24)$$

and \mathbf{p} satisfies (20). The constant vector $\boldsymbol{\gamma} \equiv -\ln(\mathbf{1} - \mathbf{c})$ in the target (22) is a function of the relative confidence in each view, as summarized by the

vector $\mathbf{c} \in (0, 1)$. The entries in γ are always positive: they are close to zero for low-confidence views and become arbitrarily large for high-confidence views. Therefore, the target (22) perturbs low-confidence views, while leaving high-confidence views untouched. In particular, notice that if the original stress-test constraints (19) are consistent, then, as expected, $\delta\mathbf{b} = \mathbf{0}$.

The optimization (20)-(22)-(23)-(24) is an instance of linear programming in the variables (\mathbf{x}, \mathbf{p}) , whose dimension equals the number of stress-tests plus the number of joint scenarios. This problem can be easily solved numerically even when such dimension is large.

In our example the adjustment is not required, i.e. $\delta\mathbf{b} = \mathbf{0}$, because the stress-tests are fully consistent with a probability set, please refer to the MATLAB code.

The posterior

Consider the prior distribution, represented by the scenarios-probabilities pair $(\mathcal{X}, \mathbf{p})$. Consider an alternative distribution on the same scenarios $(\mathcal{X}, \mathbf{q})$, where \mathbf{q} is a new vector of probabilities. We measure the proximity of the two distributions by their relative entropy

$$\mathcal{E}(\mathbf{q}, \mathbf{p}) \equiv \sum_{j=1}^J q_j \ln \frac{q_j}{p_j}. \quad (25)$$

Indeed, the relative entropy is null only when $\mathbf{q} \equiv \mathbf{p}$ and otherwise it is positive, growing larger as \mathbf{q} diverges from \mathbf{p} .

The posterior distribution is represented by new probabilities $\tilde{\mathbf{p}}$ that are as close as possible to the prior probabilities, but that reflect the stress-tests, i.e. they satisfy (19)-(21). Therefore the full-confidence posterior is defined as

$$\tilde{\mathbf{p}} = \underset{\mathbf{A}\mathbf{q} \leq \mathbf{b} + \delta\mathbf{b}}{\operatorname{argmin}} \mathcal{E}(\mathbf{q}, \mathbf{p}). \quad (26)$$

This is an instance of convex programming with linear constraints and thus in principle it is possible to solve it. However, the number of variables J , which is the number of joint scenarios, can be prohibitively large for practical purposes: for instance, for $N \equiv 10$ risk drivers and $K \equiv 3$ discrete outcomes each, we have $J = 3^{10} = 59,049$ joint scenarios, see also the case study in Section 3. Fortunately, the dual formulation of (26) is a simple convex problem in a number of variables equal to the number of views and thus it can be solved numerically very efficiently, see the proof in Meucci (2008).

In our example, we compute the probabilities (26) that minimize the relative entropy with the uniform prior (5) under the constraint (17). The result is

$$\tilde{\mathbf{p}} = \frac{1}{12} (0.9, 0.9, 1, 1, 1.05, 1.05, 1, 1, 1.05, 1.05, 1, 1), \quad (27)$$

please refer to the MATLAB code.

3 Case study: stress-testing a global market

In this section we consider a realistic case of stress-testing. Please refer to the MATLAB code for all the details and to replicate these result.

We consider a market driven $N \equiv 9$ risk factors: Z_1 is the two-year and Z_2 the ten-year points of the swap curve; Z_3 is the CDX credit default swap index; Z_4 is the S&P 500 stock market index; Z_5 is the VIX index of the implied volatility in the market; Z_6 is the dollar currency strength index; Z_7 is the crude price; Z_8 is the gold price; and Z_9 is the ten-year inflation swap rate.

We discretise our drivers into three buckets: the risk factors stay within a given range, or widen above a given threshold or widen below another threshold. More precisely, we define

$$X_n \equiv \begin{cases} 1 & \text{if } \Delta Z_n > \bar{q}_n \\ 0 & \text{if } \underline{q}_n \leq \Delta Z_n \leq \bar{q}_n \\ -1 & \text{if } \Delta Z_n < \underline{q}_n \end{cases}, \quad n = 1, \dots, N, \quad (28)$$

where ΔZ_n is the change of a risk factor over a day, and the thresholds \underline{q}_n and \bar{q}_n are the lower and upper historical terciles.

We collect all the joint scenarios for \mathbf{X} in a $J \times N$ panel \mathcal{X} , where $J = 3^9 = 19,683$ and $N = 9$.

The prior

To define the prior, we must assign a probability p_j to each of the 19,683 joint scenarios for the risk drivers, thereby obtaining the probability vector \mathbf{p} .

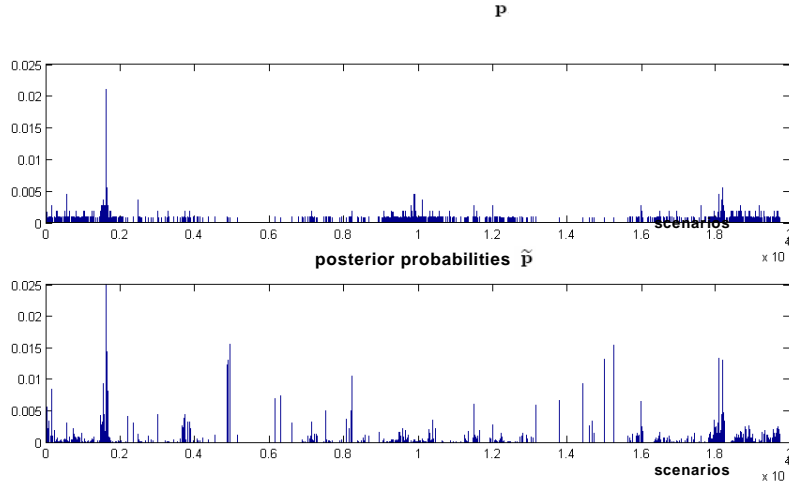


Figure 1: Prior (shrunk historical) versus posterior (stress-tested) probabilities

We do so with a frequentist estimate. First we collect the daily time series of the risk factors \mathbf{Z} from September 1, 2005 to March 30, 2010. Then we compute

the daily changes of \mathbf{Z} over the above time period and we calculate the lower and upper terciles that appear in (28).

Then we can count how many times each row in the panel \mathcal{X} was realized historically. For instance, the first row is a set of -1 , corresponding to the scenario where all the risk drivers jointly trespassed their lower tercile: such scenario never materialized and thus the frequentist probability is null.

The above process yields the purely frequentist J -dimensional vector \mathbf{p} of prior probabilities. To ensure that no scenario is strictly impossible under the prior distribution, we shrink the frequentist estimation to the uniform distribution

$$\mathbf{p} \mapsto (1 - \epsilon) \mathbf{p} + \epsilon \frac{\mathbf{1}}{J}, \quad (29)$$

where we set the shrinkage factor as $\epsilon \equiv 0.01$.

In the top plot in Figure 1 we report the bar plot of the prior probabilities (29). To gain a better understanding of such probabilities, in Table 30 we report the correlations (in %) among the risk drivers as implied by the prior distribution

	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
X_1	77	-15	24	20	4	-8	-19	29
X_2	.	-15	21	12	5	-5	-19	31
X_3	.	.	-44	10	-11	-4	38	-15
X_4	.	.	.	-9	14	5	-71	15
X_5	-29	-41	5	-16
X_6	31	-8	24
X_7	-2	11
X_8	-13

(30)

The stress-tests

We perform a stress test. First, we impose fatter tails on the scenarios, by increasing the probability of extreme events:

$$\tilde{\mathbb{P}}\{X_n = -1\} \geq 40\%, \quad \tilde{\mathbb{P}}\{X_n = 1\} \geq 40\%, \quad n = 1, \dots, N. \quad (31)$$

In other words, the probability of falling in the upper or lower tercile of all the indicators becomes at least 40%, whereas by construction historically it was 33%.

Second, we state that the probability of swap rates falling or remaining stable and gold rising, conditioned on the stock market rising is at least 90%

$$\tilde{\mathbb{P}}\{X_1 \in \{-1, 0\} \cap X_2 \in \{-1, 0\} \cap X_8 = 1 | X_4 = 1\} \geq 90\% \quad (32)$$

Using the rules (11)-(12) we convert such statements in linear constraints in the format $\mathbf{A}\tilde{\mathbf{p}} \leq \mathbf{b}$ as in (19).

Finally, we assess that the correlation between the two-year and the ten-year points of the curve will not exceed 60% over the next day, due to some awaited announcement.

$$\tilde{\text{Cor}}\{X_1, X_2\} \leq 60\%. \quad (33)$$

This statement can be added to the above in the form of a linear constraint on the posterior probabilities as in Meucci (2008), giving rise to a global set of linear constraints $\mathbf{A}\tilde{\mathbf{p}} \leq \mathbf{b}$.

Consistency check

The marginal statements (31) are fully consistent. Indeed, if we only stress-tested the system by imposing (31), the algorithm (22), would yield a null adjustment vector $\delta\mathbf{b} = (0, \dots, 0)'$ for the upper boundary \mathbf{b} in the linear constraints $\mathbf{A}\tilde{\mathbf{p}} \leq \mathbf{b}$ that represent the stress-tests.

On the other hand, the conditional statement (32) is inconsistent, although only minorly, with any probability distribution. Therefore, we relax the linear constraints $\mathbf{A}\tilde{\mathbf{p}} \leq \mathbf{b} + \delta\mathbf{b}$, where

$$\delta\mathbf{b} \approx (0, \dots, 0, 10^{-6})' \quad (34)$$

follows from (22). Notice that, as expected, only the last constraint is relaxed, namely the one corresponding to the conditional stress-test.

The statement on correlation does not violate any probability distribution, and therefore it does not elicit a perturbation of the respective upper boundary.

The posterior

With the prior probabilities (29) and the stress-tests (31)-(32) in the format of constraints $\mathbf{A}\tilde{\mathbf{p}} \leq \mathbf{b}$ we can now compute the posterior distribution as in (26). In the bottom plot in Figure 1 we report the bar plot of the posterior probabilities $\tilde{\mathbf{p}}$. To gain a better understanding of such probabilities, in Table 35 we report the correlations (in %) among the risk drivers as implied by the posterior distribution

	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
X_1	85	-34	57	13	10	-5	-52	38
X_2	.	-33	55	4	14	-3	-52	42
X_3	.	.	-54	13	-16	-4	50	-26
X_4	.	.	.	-7	19	4	-86	34
X_5	-40	-46	3	-24
X_6	35	-14	32
X_7	-1	12
X_8	-32

(35)

When the stress-test on the correlation (33) is added to the picture we obtain a new set of posterior probabilities $\tilde{\mathbf{p}}$ and a different global correlation matrix

	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
X_1	50	-26	52	16	8	-4	-45	27
X_2	.	-38	54	-6	14	-1	-55	41
X_3	.	.	-53	16	-7	-6	50	-28
X_4	.	.	.	-9	16	7	-84	38
X_5	-35	-46	6	-28
X_6	34	-14	30
X_7	-5	15
X_8	-38

(36)

Notice how the correlations in the prior (30) change to (35) as a consequence of the conditional and marginal views (31)-(32) and how they change even further to (36) due to one single statement on one correlation.

References

- Meucci, A., 2008, Fully flexible views: Theory and practice, *Risk* 21, 97–102
Extended version available at <http://ssrn.com/abstract=1213325>.
- Rebonato, R., 2010, A Bayesian approach to stress testing and scenario analysis, *Journal of Investment Management*, *Third Quarter*.