

**Simulation Engine Model**

**Prepared By: Tanya Tamarchenko**

**Date: 9/30/2014**

Contents

[Monte-Carlo Scenario Generation Engine 1](#_Toc394567744)

[Executive Summary 3](#_Toc394567745)

[Overview of Methodology 3](#_Toc394567746)

[In-Scope Products and Processes 4](#_Toc394567747)

[Model Inputs, Parameters, Assumptions and Qualitative Adjustments 4](#_Toc394567748)

[Description of Calculations 5](#_Toc394567749)

[Calibration Procedures 8](#_Toc394567750)

[Model Outputs Testing 10](#_Toc394567751)

[Model Application and Limitations 16](#_Toc394567752)

[Ongoing Performance Monitoring Approach 16](#_Toc394567753)

[Systems and IT Infrastructure 16](#_Toc394567754)

[Regulatory Requirements 16](#_Toc394567755)

[Key Personnel 16](#_Toc394567756)

[Vendor Details 17](#_Toc394567757)

[Change Log 17](#_Toc394567758)

[References 17](#_Toc394567759)

[Revision History 17](#_Toc394567760)

[Appendix A: Control Parameters 18](#_Toc394567761)

[Appendix B: BNYM Risk factors: summary as of 5/27/2014 19](#_Toc394567762)

[Appendix C: Using PCA in MC simulations for Risk Factors with missing data 20](#_Toc394567763)

[Appendix D: Lanczos subspace reduction to calculate largest eigenvalues and corresponding eigenvectors for a symmetric matrix 22](#_Toc394567764)

[Appendix E: Proxy mapping used in the current production for VaR 24](#_Toc394567765)

[Appendix F: The list of factors and their tenors for which empirical distribution is used 26](#_Toc394567766)

[Appendix G: Risk factors with no or poor quality (stale) data 28](#_Toc394567767)

# Executive Summary

The model generates Monte-Carlo (MC) simulations for all risk factors in BNYM portfolio that are used to calculate various risk measures Value-at-Risk (VaR), Stressed VaR (SVaR)). The model covers interest rates and its volatilities, equities and its volatilities, foreign exchange (FX) rates and its volatilities.

The MC simulations are calibrated to match historically observed volatilities and correlations over the period of two years for VaR and one year for SVaR. Normal Copula model is used to describe multivariate joint distribution of the risk factors. The marginal distributions for the risk factors returns are distributed normally for all factors except for the list of factors provided as an input for which the marginal distributions are generated in agreement with empirical distributions. Exponential weighting of historical returns with decay factor 0.993 is used to calculate historical volatilities and correlations.

Existing production version of MC scenarios generator has been developed over a number of years. The original developers of the code are not with BNYM anymore. The review of the model (1) resulted in a number of issues related to the PCA for equities (2) and proxies’ usage. The new MC engine eliminates PCA for the majority of equities therefore lifting the related validation issues. The lack of documentation and the fact that the original developers of the model are not available required this to be substantially rewritten. For simulating the risk factors other than equities prices there is no change to the methodology per se, the main features of the model are unchanged. The mechanics of the calculations are made simpler, a few unnecessary steps are eliminated, and the number of reports is generated to make the process more transparent. The MC scenario generation (the production and the new version as well) runs independently from RiskWatch. This development also made the links to RiskWatch (mainly naming conventions for the Risk Factors) more transparent.

# Overview of Methodology

The proposed new approach will lead to:

* Elimination of principal component analysis (PCA) for equities that have sufficient historical data;
* Elimination of Cholesky (or SVD) in scenario generator;
* Elimination of idiosyncratic returns simulation for equities with no substantial gaps in historical time series;
* Elimination of the need to calculate and store variance-covariance or correlation matrix across all the factors;
* Reduction of the number of independent driving risk factors from about 2000 to less than 500.

The main features of current production MC scenario generation model remain intact:

* Generated simulations are designed to replicate historically observed variance-covariance matrix across the risk factors;
* Exponential weighting is used;
* Empirical marginal distribution for the risk factors is used for the list of the risk factors specified in the input file;
* The proxyying is used for the risk factors in the specified input file, i.e. the simulated returns of a proxied risk factor are generated as perfectly correlated with the simulated returns of the proxy factor.

The PCA process still has a limited use, specifically to handle equities with missing data. The numerical procedure to implement PCA is reworked and does not require random sub-matrix selection anymore.

# In-Scope Products and Processes

The output of the model (MC scenarios) is passed to ALGO which revaluates the BNYM trades and produces VaR and SVaR numbers.

# Model Inputs, Parameters, Assumptions and Qualitative Adjustments

The input data for each date include the following:

*Historical Data:*

hr\_spot.csv

hr\_curve.csv

hr\_spot\_<date>\_All.csv

*End of date (EOD) data (usage is not currently implemented)*

fx\_spot.csv

fx\_zerocurves.csv

indexCurves.csv

ir\_zerocurves.csv

other\_zerocurves.csv

MBS\_EOD\_Values.csv

creditbond\_zerocurves.csv

fx\_atmvols\_0.csv

*The lists of previous day risk factor names*:

eqtRWHOpenPositionReport.csv

riskFactorTemplateBNYE.csv

riskFactorTemplateBNYMCMNOSPRISK.csv

riskFactorTemplateBNYMCMSPRISK.csv

riskFactorTemplateCMNOSR.csv

riskFactorTemplateCMSR.csv

riskFactorTemplateCstIR.csv

riskFactorTemplateEQUITY.csv

riskFactorTemplateEQUITYSR.csv

riskFactorTemplateFX.csv

riskFactorTemplateFXNEW.csv

riskFactorTemplateIR.csv

riskFactorTemplatePERSHING.csv

riskFactorTemplatePERSHINGNOSR.csv

riskFactorTemplatePERSHING.SRcsv

*Proxy Maps*

ProxyList.csv (used for SVaR)

RWLink\_Proxies.csv ( used for VaR and SVaR)

*Control Parameters:*

control\_parameters.csv includes the following parameters, see more details in Appendix A.

*The list of factors for which empirical distribution is used:*

realFactors.csv, see Appendix F.

# Description of Calculations

1. From the input files with the list of yesterdays’ positions risk factors create a list of unique factor names. For these RF the simulations will be created.
2. Collect historical data for equities, equity volatilities and all other risk factors in the same manner as it is done presently.
3. Identify risk factors with low number of missing data (Set 1, includes N risk factors) and the rest that have significant gaps in available historical data (Set 2 includes M risk factors). This step is in place in current system.

Steps 3 – 7 are performed for Set 1

1. Calculate log-returns (or differences) over the specified time horizon;
2. Identify set of dates for which historical data is available across all risk factors. K is the number of such dates. K should be between 480 and 500 for VAR. This step is in place in current system;
3. Form a N x (K-1) matrix R of returns;
4. Calculate K-1 weights based on decay factor (d=0.993 in production now). This is done in current system. Form a diagonal matrix of weights W.
5. Generate each Monte-Carlo (MC) scenario as follows:

* Generate vector v of K independent random normal variables;
* Calculate simulated returns by multiplying matrix R by vector v:

(1)

*Theoretical justification:*

Variance-covariance matrix (VCV) of simulated returns is:

(2)

Note that the simulations are generated for all equities that are part of Set 1. The change to the dlm in RiskWatch was implemented by Tielong Xie to handle the equities using their simulations when they are available instead of using the factor coefficients for the principal components.

1. Group the RF from Set 1 into sub-groups according to the risk factor type (Equity Spot, Equity Vol, FX Spot, FX Forward, IR Curve, IR Vol, FX pairs vol and perform PCA for each sub-group. The number of PC is specified in the input file and is currently set to 18 to be consistent with current production. The detailed description of the PC and its use in simulations is in Appendix C.
2. Use the PCA analysis to handle the RF with missing data to perform regression.
3. Generate simulations for the RF from Set 1 and then for the rest using the regression coefficients. Transform simulations to empirical distribution for the RF assigned to be simulated according to empirical distribution.
4. In order to handle the equities for the Pershing positions that are not yet known at the time of the scenario generation run, output the equity PC historical data. This part is similar to the current production run of the “gbFactor” module.
5. Generate simulations for the factors lacking historical data and identified as “idiosyncratic”.
6. Generate a scenario file.

Currently the list of BNYM risk factors includes about N = 4500-5000 factors, see the summary in the Appendix B. Note that the rank of the VCV does not exceed the number of columns if the matrix R in (1) which equals to the number of historical returns K. This number K is less than 500 for VaR and 250 Stressed VaR. As can be seen from (1) we need just K independent random variables to simulate any number N of risk factors to reproduce the historical VCV based on these K returns. Note that increasing the number of risk factors while keeping unchanged the number of MC simulations and the length of the historical data period upon which the calibration is based, does not affect the variance of the VaR estimate with the order statistics. This can be seen from equation (1) above that shows that all simulations are generated with the number of independent normal variables equal to the number of historical returns used for calibration.

With the proposed approach we eliminated a few unnecessary extra steps performed by the legacy model for scenario generations, namely:

- VCV calculation;

- PCA decomposition for equities for the equities in the “morning” batch;

- Cholesky decomposition for all risk factors.

We also reduced the number of independent random variables to K (not exceeding 500 now) versus about 2000 in the current production system.

*Dealing with missing or stale data*

Theoretically the described model is straightforward. In practice however it becomes much less straightforward due to the fact that historical data for some factors are not available for some dates or stale. The current production module has some attempt of ad hoc data completion procedure. In the new version we tried to avoid implementing ad hoc solutions and deal with insufficient data by calculating principal components for each risk factor type and performing regression for each RF with insufficient data.

The Table 1 below presents the summary of the risk factors with no gaps in the historical two year period preceding the VaR run date 5/27/2014.

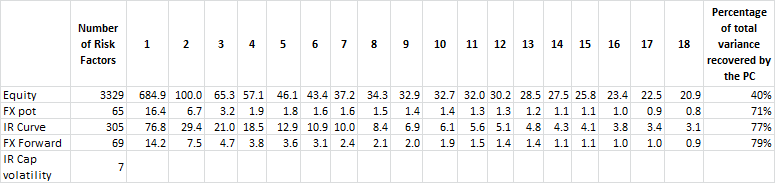
Table 1: Risk factors with no gaps in the historical data

|  |  |
| --- | --- |
| **Risk Factor Type** | **Count** |
| **Total** | **3537** |
| Equity Volatility | 75 |
| Equity Spot | 3329 |
| FX Forward | 17 |
| FX Pair Volatility | 5 |
| FX Spot | 65 |
| IR Curve | 37 |
| IR Cap Volatility | 7 |
| MBS | 1 |
| CDS | 1 |

The risk factors with no or stale data are listed by type in the Appendix G. The list of factors with some data available is also shown in the Appendix G together with the treatment in simulations (“Regressed” or “Idiosyncratic”). This way of dealing with missing data allows for the transparency with respect to the data quality of the historical data and its’ effect on the simulations.

The risk factors that had fewer than 21 days of good quality historical data (no gaps), where simulated as idiosyncratic with default volatility.

The rest were regressed against the principal components for the corresponding risk factor type. The number of principal components to be used for the regression is specified in the input control file. Currently it is set to 18 for all types. The principal components for equities spot prices are shown in the Appendix D. The resulting eigenvalues and the percentage of variance recovered with these PC is presented in the table 2 below. The IR Cap volatility factor type comprises of only 7 risk factors: 6 caplet volatilities for USD and 1 for EUR, therefore PCA was not performed for this risk factor type and the historical returns for these factors themselves would be used for regression. In fact all factors for this type had good quality historical data, so no regression was performed with these.

Table 2: PCA results for different risk factor types

*Using proxies*

The current production version uses “proxies”, i.e. mapping of the risk factor. The content of the file RWLink\_Proxies.csv is shown in the Appendix E. The mappings specified in this file mean that the simulated return of the mapped factors is perfectly correlated with the simulated return of the factor it mapped to. The number of FX Forward curves (interest rate curves in different currencies) is mapped to the FX spot risk factors (spot exchange rate). In addition different interest rate curves in each currency are typically mapped to a single Libor curve in this currency. For example in USD such curves as FEDFD, LIB1Y, LIB6M, TBILL, MUNI are mapped to the Interbank curve which is LIB3M. The new model currently does use the same mappings for the purpose of comparing the results to the current production.

*Simulating log-returns versus differences*

For the majority of the RF we implicitly assume that the distribution of their log-returns is stationary. Therefore we use their log-returns in the statistical analysis of the historical data. Currently only a single risk factor is an exception – this is USD\_MBS\_CC\_OAS which represents the spread and can take negative values. It is not possible to deal with this factor via the log-returns of its historical values. Therefore differences are used instead of log-returns.

*Using empirical distribution instead of normal distribution for simulated log-returns and differences*

In the current production some tenors for a specified list of risk factors can be simulated using their empirical distributions. For the rest the log-returns or differences are simulated as normal. The new version also implements this approach. The list of the risk factors and their tenors for which empirical distribution is used is given in Appendix F.

# Calibration Procedures

Monte-Carlo simulations are calibrated to historical period of most recent two years of data for VaR and one year stressful period in the past. In order to generate scenarios for VaR and Stressed VaR the risk factors (RF) underlying BNYM portfolio on a given day are grouped in 3 sets:

* Set 1: RF with no or few days of missing data;
* Set 2: RF with more than few dates of missing data but still enough data to estimate regression;
* Set 3: RF with no data or few days of data not enough to do regression.

In this implementation we intend to avoid time series completion procedures for the RF with missing data. The input parameter minDatesCount controls the number of dates that each RF from set 1 should have. The RFs from Set 1 are dealt with via the following steps:

* The RF that have more than minDatesCount “good quality data” are assigned as member of Set 1;
* The common dates where historical data for RF in Set 1 are available, are identified.
* The matrix of log-returns is calculated based on these common dates;
* The Principal Component Analysis (PCA) is performed for each risk factor type. This is only done to handle RF from Set and is described in more detail below. Historical time series corresponding to the PC are calculated;
* Independent normally distributed standard normal variables are generated;
* Simulated returns are calculated according to equation (1) above;
* The simulated returns are transformed to comply with empirically observed distributions for each RF.

To identify “good quality data” some data quality control is implemented. The input parameters maxAllowedGap, maxDatesUnch and staleDataDailyVolMin control this process. If the data has gaps exceeding maxAllowedGap or stretches of unchanging data exceeding maxDatesUnch it daily volatility below staleDataDailyVolMin, this data is considered low quality and eliminated. Some factors from Set 1 with low quality data might be assigned to Set 2 or Set 3.

In order to deal with the RF from Set 2 we employ regression analysis of these risk factors against principal components (PC) for the corresponding risk factor type. To achieve this we perform Principal Component Analysis (PCA) for the RF from Set 1 for each of the following risk factor types:

* Equity Spot;
* Equity Volatility;
* FX Forward;
* FX Pair Volatility;
* FX Spot;
* IR Curve;
* IR Cap Volatility.

Then each RF from the Set2 is regressed against appropriate set of PCA. The resulting regression coefficients and residual volatility are used to generate simulation. These calculations are described in more details Appendix C. The numerical algorithm for PCA based on Lanczos subspace reduction (3) is described in Appendix D.

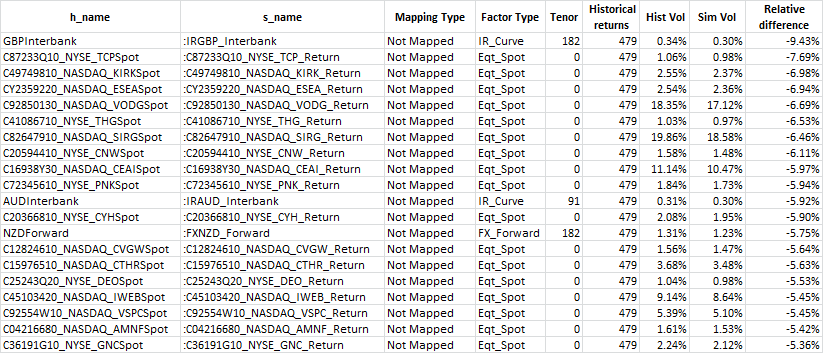
# Model Outputs Testing

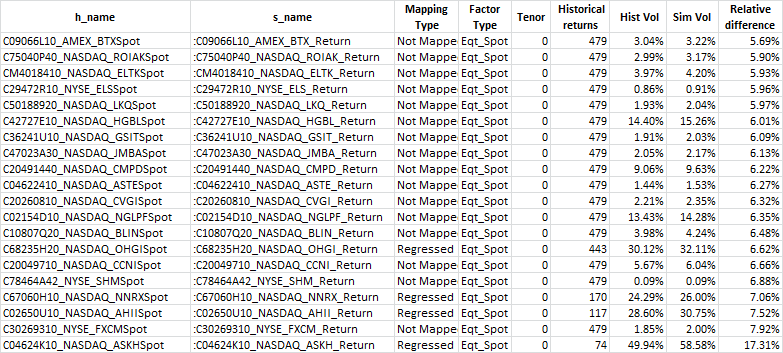
Unless it is noted otherwise, all test results described below where obtained for the VaR run as of 5/27/2014 based on 2 years of historical data. The number of MC simulations is set 1000.

*Comparing volatilities of simulated risk factors to the historical volatilities*

The MC simulations are designed to replicate the historically observed volatilities and correlations across the risk factors. The good agreement between observed and simulated numbers is expected for all risk factors that were not proxied and simulated as log-normal. The table 4 below presents the results for 40 risk factors with the largest relative differences between historical volatility and the volatility of the simulations. The relative differences between the volatilities for the rest 4328 risk factors are between -5.3% and 5.7%, for 3800 of them the relative difference does not exceed 3%.

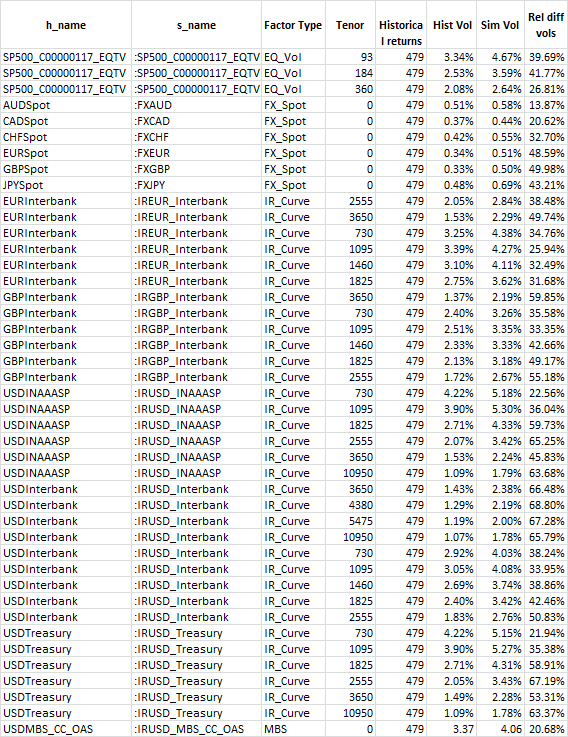
Table 4: Simulated versus historical volatilities for the factors that are not proxied and simulated as log-normal.





The Table 5 below shows the comparison between the factors with empirical distributions As expected the volatilities of the simulated RF are 30%-60% higher that the observed volatilities.

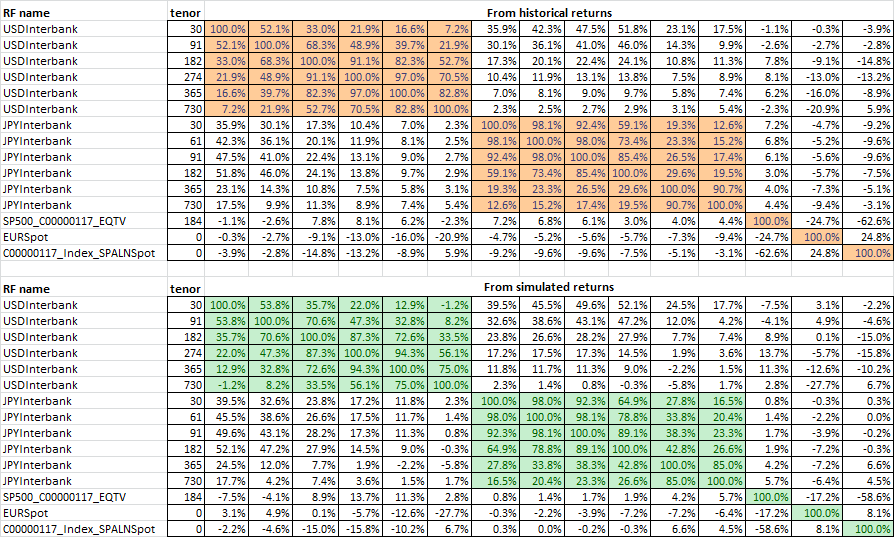
Table 5: Risk factors with empirical marginal distribution: comparison volatilities



*Comparing correlations of simulated risk factors to the historical correlations*

The comparison of the correlations for few major risk factors is shown in the Table 6. The historical and simulated correlations are similar.

Table 6: Correlations comparison for major risk factors

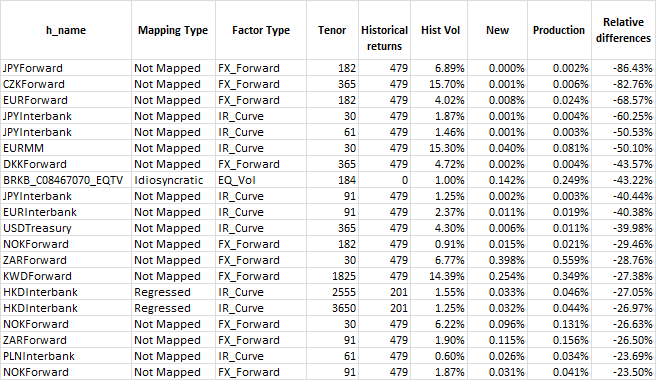


*Comparing New simulations to Production*

In this section we examine the discrepancies in terms of standard deviation of simulated daily differences between production and the new simulations. The Table 7 presents the results for the factors simulated as log-normal, without proxyying for the 40 RF with the largest discrepancies. It was established in the section above that the new simulations are in a very good agreement with the historical data. The cases with the largest discrepancies might be investigated further.

The comparison between the new and production simulations for the factors with empirical distribution is illustrated in the Table 8. For all risk factors in this group the new simulations have higher standard deviation compared to the production by about 20%. It looks as if the transformation to the empirical distribution was not applied in the production.

Table 7: New vs production simulations: factors with log-normal distribution



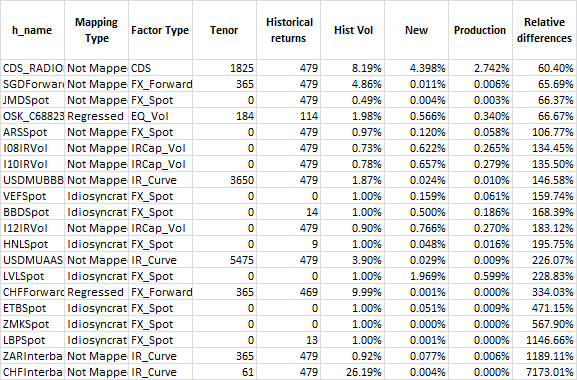
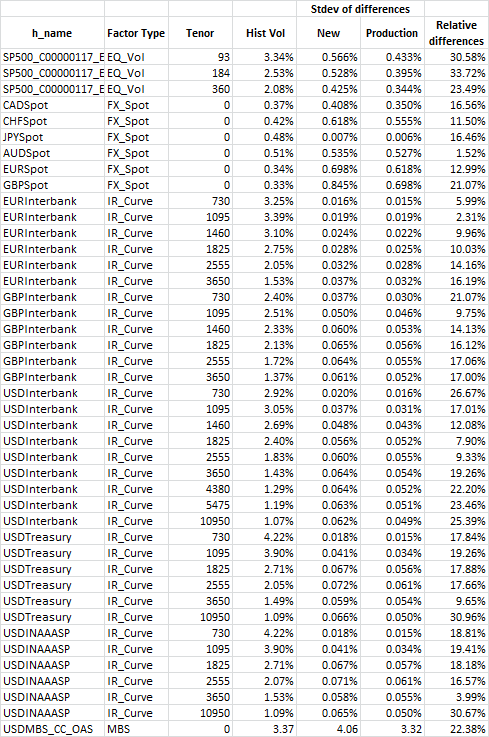


Table 8: New vs production simulations; factors with empirical distribution

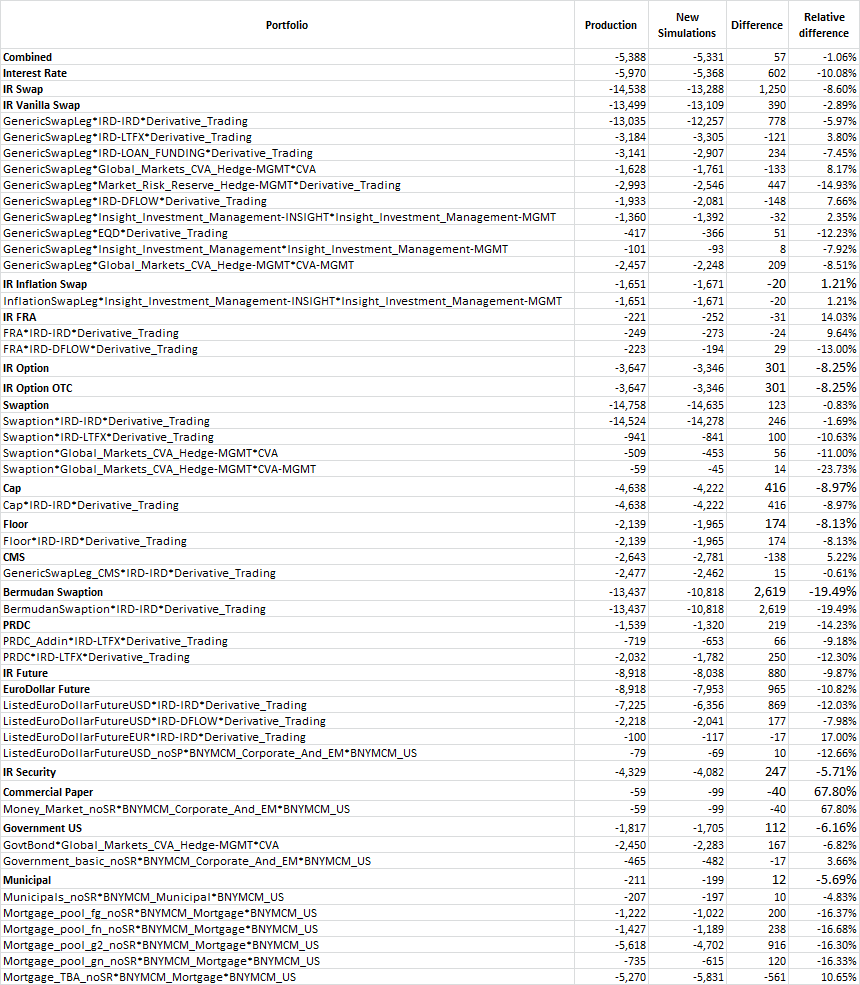


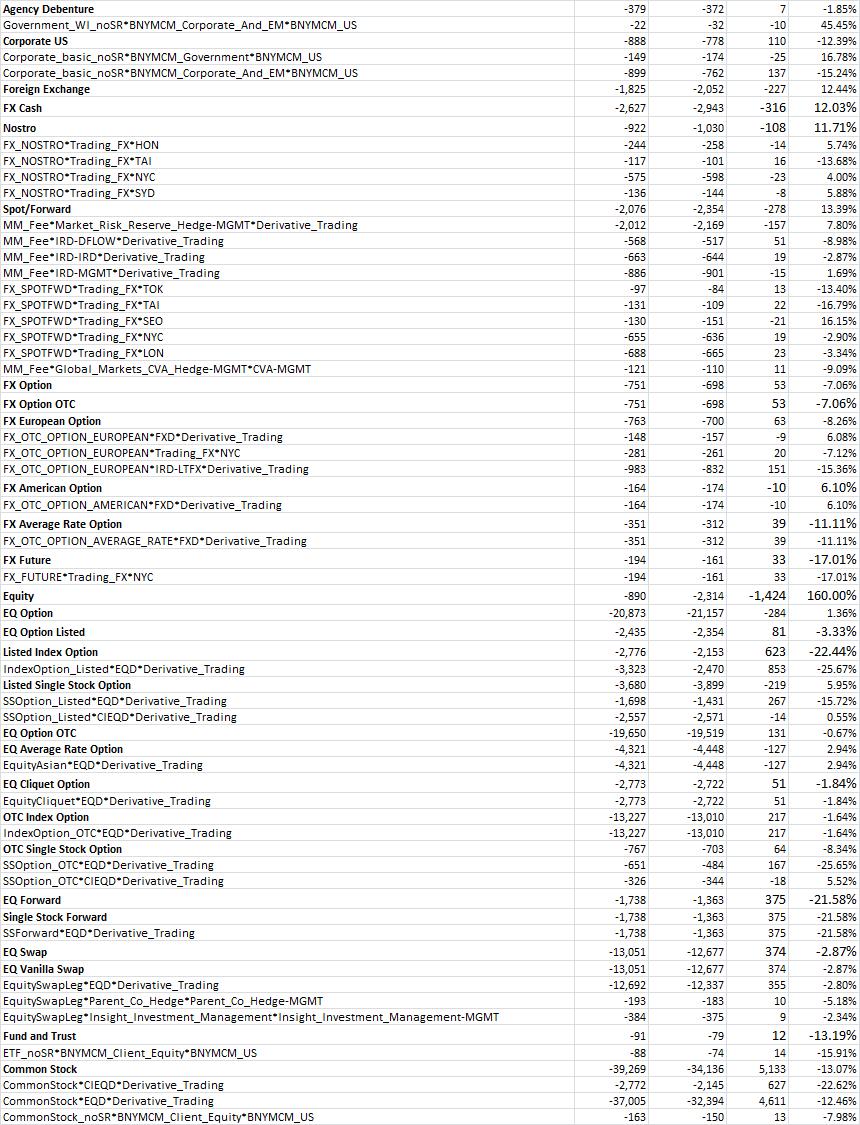
*New simulations effect of VaR*

The new simulations file was passed to ALGO which generates VaR reports for BNYM portfolios. We conducted 2 types of tests. Firstly, we generated simulations in the new engine by simulating all equity price risk factors via Principal Components in the manner that is identical to what is used currently in Production. The call this test “New Simulations Effect”. The results are presented in the Table 9 below. In the second test we treated equity simulation as it would be after the new engine is released into production. In production the simulation engine for the current date is run before the information regarding all current position is available. For the equities that are not yet known at the time of the scenario generation run, simulation is performed via PCA. We call this test “Equity Remediation Effect”. This test results are shown in the table 10. The tables 9 and 10 present all the portfolios for which the difference in VaR numbers exceed 7K. The tests are run as of 8/19/2014.

Table 9: Production vs New Simulations: VaRs comparison: new simulations effect

Table 10: Production vs New Simulations: VaRs comparison: equity remediation effect





*Informational reports*

The reports generated by the new model show the list of risk factors by risk factor type plus detailed information relevant to scenario generation including the number of historical samples with good quality data used in calibration, historical volatility, the way the factor was treated in simulations (not mapped, regressed, idiosyncratic) plus the 1000 simulated log-returns (see example in the Table 4 above). This type of information included in the reports helps to identify and explain the VaR numbers resulting from the simulations.

# Model Application and Limitations

The Phase 1 of the implementation includes generation scenarios for all BNYM factors, passing this scenario file to ALGO in the same manner it is done for the production version of the scenarios. The model includes the following features:

* Exponential weighting of the historical returns;
* Handling RF with incomplete , missing or stale historical data;
* Proxy mapping for the specified subset of factors;
* Using empirical distribution for the specified subset of factors;
* Generating simulations for all equities for which historical data is available, without performing any PC analysis according to historically implied correlations;
* Elimination a number of calculations performed by the current production code, specifically:

∙ Computing variance-covariance matrix across the RF;

∙ Performing Cholesky decomposition;

∙ Using PCA to simulate all equities

* The PCA analysis has limited use for the RF with incomplete of stale historical data;
* The algorithm of computing PC is reworked and does not use random sub-matrices anymore.
* The generated set of MC scenarios includes the scenarios of all equities which are used by RiskWatch in similar manner as the simulations for all other risk factors to shift equities prices. Factor coefficients are not needed anymore.
* The detailed reports are generated by the model that allow for easier analysis of the results .

# Ongoing Performance Monitoring Approach

The ongoing performance monitoring that is already in place will be applicable after the new simulations go into production. This includes back testing procedures and monitoring,

# Systems and IT Infrastructure

The new model runs in Linux environment. There is no change compared to the current production model.

# Regulatory Requirements

Document here the model’s compliance with any applicable regulatory rules (e.g. Fed, SEC, FSA) and agreements with securities exchanges.

# Key Personnel

Model owner: Tanya Tamarchenko

Model developer: Tanya Tamarchenko

Model user: Market Risk

Model control personnel: ?

# Vendor Details

N/A

# Change Log

Phase 1: 7/30/2014

# References

1. Model Risk Management Group, Market Risk VaR validation report, 6/27/2013.
2. Model Risk Management Group, Appendix E - Assessment of Equity PCA Methodology, 6/27/2013.
3. C.C. Paige, Computational Variants of the Lanczos Method for the Eigenproblem, Journal of the Institute of Mathematics and its Applications, 10, 373-381, 1972.

# Revision History

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **DATE** | **SECTION** | **BRIEF DESCRIPTION** | **VERSION #** | **REQUESTED BY** |
| 7/22/14 | All | Development Phase 1 | 1.0 | ERM |
|  |  |  |  |  |

# Appendix A: Control Parameters

**Parameter, Value, Comment**

minDatesCount, 480, Minimum number of dates a factor should have to avoid being regressed

num\_pca\_ir\_curve, 18, Number of principal components calculated for IR to use in regressions

num\_pca\_fx\_forward, 18, Number of principal components calculated for FX Forward to use in regressions

num\_pca\_eqt\_spot, 18, Number of principal components calculated for Eqt Spotto use in regressions

num\_pca\_eqt\_vol, 18, Number of principal components calculated for Eqt Vol to use in regressions

num\_pca\_fx\_spot, 18, Number of principal components calculated for FX Spot to use in regressions

num\_pca\_fx\_vol, 18, Number of principal components calculated for FX Vol to use in regressions

minNumRetsForRegr, 21, Minimum days of data without gap a factor should have to be regressed

maxAllowedGap, 5, Maximum allowed gap in the data. If a gap exceeds this number, all previous values are thrown away

maxDatesUnch, 10, Maximum allowed number of dates with unchanged data. If a number of days with unchanged data exceeds this number, all previous values are thrown away"

defaultIdioVol, 0.01,

staleDataDailyVolMin, 0.0,

decay\_factor, 0.993,

percentile, 0.01

randomness\_control, 0, Controls randomness

largest\_return, 0.5, Largest allowed return in historical data

shift, 0.0001, Shift to set scenarios for sensitivities

sim\_horizon, 1, Simulation horizon in days

# Appendix B: BNYM Risk factors: summary as of 5/27/2014

Note that cure data are bucketed so that one Interest Rate, FX forward curve, or Equity Volatility term structure is normally represented by a few tenors (a few risk factors)

Table A1: The BNYM Risk Factors summary by type

|  |  |
| --- | --- |
| **Type** | **Count** |
| **Total** | **4699** |
| Equity Vol | 108 |
| Equity Spot | 3813 |
| FX Forward | 80 |
| FX Pair Vol | 6 |
| FX Spot | 131 |
| IR Curve | 544 |
| IR Cap Vol | 7 |
| MBS | 3 |
| CDS | 7 |

# Appendix C: Using PCA in MC simulations for Risk Factors with missing data

Let’s consider a set of RF with good historical data. First PCs are calculated from the historical correlation matrix of the RFs returns for the set.

*Notations:*

R – N x K matrix of returns over K + 1 common dates across N risk factors;

W – diagonal K X K matrix of exponential weights;

VCV = R ∙ W ∙ RT - VCV from historical returns; (C1)

Ϭ2 = diag(VCV) - diagonal matrix of variances for the RF;

Corr = Ϭ-1 ∙ R ∙ W ∙ RT ∙ Ϭ-1 – correlation;

E – unitary matrix;

PCA of the Correlation matrix would result in the following representation (Q – orthonormal eigenvectors, i.e. QT ∙ Q = E, Λ – diagonal matrix of the eigenvalues):

Corr = Q ∙ Λ ∙ QT , (C2)

Selecting p eigenvectors corresponding to the largest eigenvalues we get the matrices Corr\* and VCV\* which are different from the original correlation and variance-covariance matrices:

Corr\* = Qselect ∙ Λselect ∙ Qselect T (3C)

VCV\* = Ϭ ∙ Qselect ∙ Λselect ∙ Qselect T ∙ Ϭ (4C)

Historical returns for p principal components (K X p):

H = W0.5 ∙RT∙ Ϭ-1 ∙ Qselect∙ Λselect-0.5 (5C)

Projection of historical RF returns on the PC (N X p matrix P):

P = R ∙ W0.5 ∙ H = Ϭ ∙ Qselect∙ Λselect0.5 (6C)

**Check: H is orthonormal**

HT ∙ H = Λselect-0.5 ∙ QselectT ∙ (Ϭ-1 ∙ R ∙ W ∙ RT ∙ Ϭ-1) ∙ Qselect∙ Λselect-0.5 = (7C)

Λselect-0.5 ∙ QselectT ∙ (Q ∙ Λ ∙ QT) ∙ Qselect∙ Λselect-0.5 = E

**Check: P x PT = VCV\***

P ∙ PT = Ϭ ∙ Qselect∙ Λselect∙ QTselect∙Ϭ = VCV\* (8C)

**In simulations:**

η = P ∙ ξ, where ξ is independent normal (9C)

Formula (9C) shows how to reduce dimensionality of the problem and simulate N factors using p independent normal variables. We don’t intend to use (9C) in the current implementation of MC for two reasons: 1) The dimensionality of our problem is limited by how many days of history we use to calibrate the model (500 for VaR and 250 for SVaR); and 2) Reducing dimensionality of the problem leads to variance-covariance matrix of the simulated RF be different from the original historical VCV and that raises the question regarding how many principal components “is enough” the answer to which depends of the portfolio construction.

The analysis above is useful however for handling the RF with incomplete data via regression analysis against the PCs for the corresponding risk factor type. We suggest using p vectors (the columns of matrix H calculated according to 5C) as independent variables. Taking a RF with incomplete data we regress it’s returns against the p independent variables (principal components). The regression will produce the vector a of p regression coefficients and the residual volatility b. The simulations for the risk factor will be calculated as follows:

where the last term represents simulated idiosyncratic risk. This approach is actually similar to what is done currently in the production for all equities.

# Appendix D: Lanczos subspace reduction to calculate largest eigenvalues and corresponding eigenvectors for a symmetric matrix

The idea of Lanczos subspace reduction for a symmetric linear operator A (correlation matrix in our application) is to build a linear space spanned on the vectors <q, Aq, A2q, …, Anq> that represent powers of the operators A applied to the initial vector q. This space can be built by means of simple three-term recursion:

, (D1)

Where:

q1 - is arbitrary vector with norm equal to 1;

(D2)

,

It is easy to verify that vectors qi form orthonormal basis. The coefficients α and β form a tri-diagonal symmetric matrix whose spectrum approximates the spectrum of the operator A. Theoretically continuing to increase the dimensionality of Krylov space iteratively via process (D2) will result in better approximation of spectrum of operator A. As shown in the relevant publications (see (3)) the largest eigenvalues and eigenvalues of matrix A can be found after few steps of iterative process (D2). Practical problem is the loss of orthogonality of vectors qi due to computational round-off error. This problem, however, has been overcome by the practitioners by performing re-orthoginalization of the vectors qi.

Note that each step of Lanczos recursion requires multiplication of the matrix A by vector q. In our case matrix A is the correlation matrix calculated from weighted returns via formula (C1) in the Appendix C above:

A = Corr = Ϭ-1 ∙ R ∙ W ∙ RT ∙ Ϭ-1

Matrix A can be written at:

A = P ∙ PT, Where PT = Ϭ-1 ∙ R ∙ W0.5. Thus we don’t need to calculate and store a large correlation matrix (about 4700 X 4700 in or case). Instead we need to use scaled matrix of returns P (about 4700 X 500 in our case) and multiply it’s transposed and itself by a vector on each step of the recursion (D1) above.

The Figure D1 below shows the PC calculated from the equity data, using 2 years of data prior to the VaR calculation date 5/27/2014. The Correlation matrix of the returns of the principal components is presented on Figure D2.

Figure D1: 18 Principal Components from Equity data

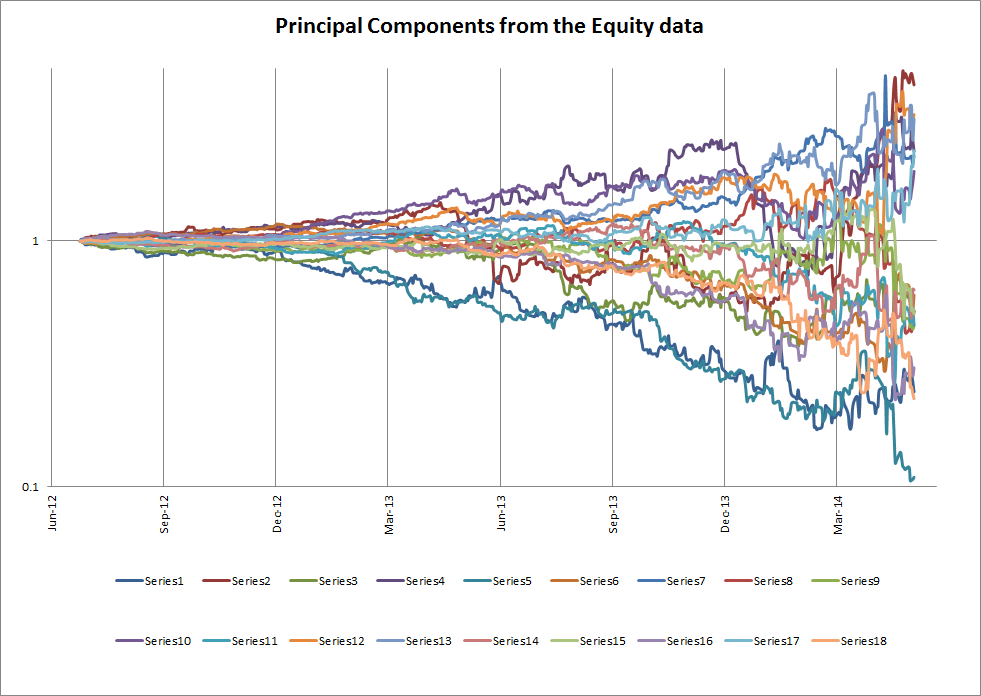
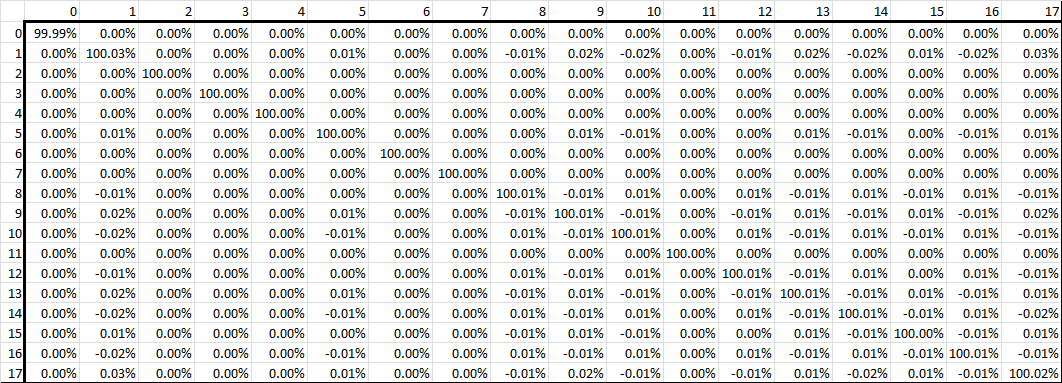


Figure D2: Variance –covariance matrix of Principal Components returns



# Appendix E: Proxy mapping used in the current production for VaR

**Risk Factor Type, Risk Factor Name, Term, Proxy Name, Proxy Multiplier, Proxy Spread, Comments**

IR, FXARS\_Forward,,ARSSpot, ,,,

IR, FXBBD\_Forward,,BBDSpot, ,,,

IR, FXBDT\_Forward,,BDTSpot, ,,,

IR, FXBIF\_Forward,,BIFSpot, ,,,

IR, FXBMD\_Forward,,BMDSpot, ,,,

IR, FXCRC\_Forward,,CRCSpot, ,,,

IR, FXDJF\_Forward,,DJFSpot, ,,,

IR, FXDZD\_Forward,,DZDSpot, ,,,

IR, FXEEK\_Forward,,EEKSpot, ,,,

IR, FXETB\_Forward,,ETBSpot, ,,,

IR, FXFJD\_Forward,,FJDSpot, ,,,

IR, FXGHS\_Forward,,GHSSpot, ,,,

IR, FXGMD\_Forward,,GMDSpot, ,,,

IR, FXGYD\_Forward,,GYDSpot, ,,,

IR, FXHNL\_Forward,,HNLSpot, ,,,

IR, FXIQD\_Forward,,IQDSpot, ,,,

IR, FXIRR\_Forward,,IRRSpot, ,,,

IR, FXJMD\_Forward,,JMDSpot, ,,,

IR, FXLBP\_Forward,,LBPSpot, ,,,

IR, FXLVL\_Forward,,LVLSpot, ,,,

IR, FXMRO\_Forward,,MROSpot, ,,,

IR, FXMWK\_Forward,,MWKSpot, ,,,

IR, FXNAD\_Forward,,NADSpot, ,,,

IR, FXPYG\_Forward,,PYGSpot, ,,,

IR, FXSCR\_Forward,,SCRSpot, ,,,

IR, FXSDG\_Forward,,SDGSpot, ,,,

IR, FXSZL\_Forward,,SZLSpot, ,,,

IR, FXTND\_Forward,,TNDSpot, ,,,

IR, FXTZS\_Forward,,TZSSpot, ,,,

IR, FXUAH\_Forward,,UAHSpot, ,,,

IR, FXUGX\_Forward,,UGXSpot, ,,,

IR, FXUSD\_Forward,,IRUSD\_Interbank, ,,,

IR, FXUYU\_Forward,,UYUSpot, ,,,

IR, FXVEF\_Forward,,VEFSpot, ,,,

IR, FXXAF\_Forward,,XAFSpot, ,,,

IR, FXXOF\_Forward,,XAFSpot, ,,,

IR, FXZMK\_Forward,,ZMKSpot, ,,,

IR, FXZMW\_Forward,,ZMWSpot, ,,,

IR, IRAUD\_BBS3M,,IRAUD\_Interbank, ,,,

IR, IRAUD\_USDCC,,IRAUD\_Interbank, ,,,

IR, IRCAD\_BA,,IRCAD\_Interbank, ,,,

IR, IRCAD\_USDCC,,IRCAD\_Interbank, ,,,

IR, IRCHF\_USDCC,,IRCHF\_Interbank, ,,,

IR, IRCZK\_PRIBOR,,IRCZK\_Interbank, ,,,

IR, IRDEM\_USDCC,,IRDEM\_Interbank, ,,,

IR, IREUR\_EIB1M,,IREUR\_Interbank, ,,,

IR, IREUR\_EIB1Y,,IREUR\_Interbank, ,,,

IR, IREUR\_EIB3M,,IREUR\_Interbank, ,,,

IR, IREUR\_EIBOR,,IREUR\_Interbank, ,,,

IR, IREUR\_EONIA,,IREUR\_Interbank, ,,,

IR, IREUR\_FRCPI,,IREUR\_Interbank, ,,,

IR, IREUR\_HICP,,IREUR\_Interbank, ,,,

IR, IREUR\_USDCC,,IREUR\_Interbank, ,,,

IR, IRGBP\_LIB1M,,IRGBP\_Interbank, ,,,

IR, IRGBP\_LIB3M,,IRGBP\_Interbank, ,,,

IR, IRGBP\_SONIA,,IRGBP\_Interbank, ,,,

IR, IRGBP\_USDCC,,IRGBP\_Interbank, ,,,

IR, IRHKD\_USDCC,,IRHKD\_Interbank, ,,,

IR, IRJPY\_LIB1M,,IRJPY\_Interbank, ,,,

IR, IRJPY\_LIB1Y,,IRJPY\_Interbank, ,,,

IR, IRJPY\_LIB3M,,IRJPY\_Interbank, ,,,

IR, IRJPY\_TIB1M,,IRJPY\_Interbank, ,,,

IR, IRJPY\_TIB3M,,IRJPY\_Interbank, ,,,

IR, IRJPY\_TIBOR,,IRJPY\_Interbank, ,,,

IR, IRJPY\_Treasury,,IRJPY\_Interbank, ,,,

IR, IRJPY\_USDCC,,IRJPY\_Interbank, ,,,

IR, IRNLG\_USDCC,,IRNLG\_Interbank, ,,,

IR, IRPLN\_WIBOR,,IRPLN\_Interbank, ,,,

IR, IRSEK\_STIBOR,,IRSEK\_Interbank, ,,,

IR, IRUSD\_AAGIC,,IRUSD\_Interbank, ,,,

IR, IRUSD\_FEDFD,,IRUSD\_Interbank, ,,,

IR, IRUSD\_LIA6M,,IRUSD\_Interbank, ,,,

IR, IRUSD\_Interbank1,,IRUSD\_Interbank, ,,,

IR, IRUSD\_LIB1Y,,IRUSD\_Interbank, ,,,

IR, IRUSD\_LIB6M,,IRUSD\_Interbank, ,,,

IR, IRUSD\_MERID,,IRUSD\_Interbank, ,,,

IR, IRUSD\_MUNI,,IRUSD\_Interbank, ,,,

IR, IRUSD\_TBILL,,IRUSD\_Treasury, ,,,

IR, IRUSD\_INAAA,,IRUSD\_INAAASP, ,,,

IR, IRUSD\_CAProvince,,IRUSD\_INAAASP, ,,,

IR, IRUSD\_Supranational,,IRUSD\_Treasury, ,,,

IR, IRUSD\_MUAAA,,IRUSD\_MUAAASP, ,,,

# Appendix F: The list of factors and their tenors for which empirical distribution is used

**RF Name, Term**

USDInterbank,730

USDInterbank,1095

USDInterbank,1460

USDInterbank,1825

USDInterbank,2555

USDInterbank,3650

USDInterbank,4380

USDInterbank,5475

USDInterbank,10950

USDTreasury,730

USDTreasury,1095

USDTreasury,1825

USDTreasury,2555

USDTreasury,3650

USDTreasury,10950

EURInterbank,730

EURInterbank,1095

EURInterbank,1460

EURInterbank,1825

EURInterbank,2555

EURInterbank,3650

GBPInterbank,730

GBPInterbank,1095

GBPInterbank,1460

GBPInterbank,1825

GBPInterbank,2555

GBPInterbank,3650

EURSpot,0

JPYSpot,0

GBPSpot,0

CHFSpot,0

CADSpot,0

AUDSpot,0

J00IRVol,0

J01IRVol,0

J02IRVol,0

J03IRVol,0

J04IRVol,0

J05IRVol,0

J06IRVol,0

J07IRVol,0

J08IRVol,0

J09IRVol,0

J10IRVol,0

J11IRVol,0

J12IRVol,0

J13IRVol,0

J14IRVol,0

J15IRVol,0

J16IRVol,0

J17IRVol,0

USDINAAASP,730

USDINAAASP,1095

USDINAAASP,1825

USDINAAASP,2555

USDINAAASP,3650

USDINAAASP,10950

USDMBS\_CC\_OAS,0

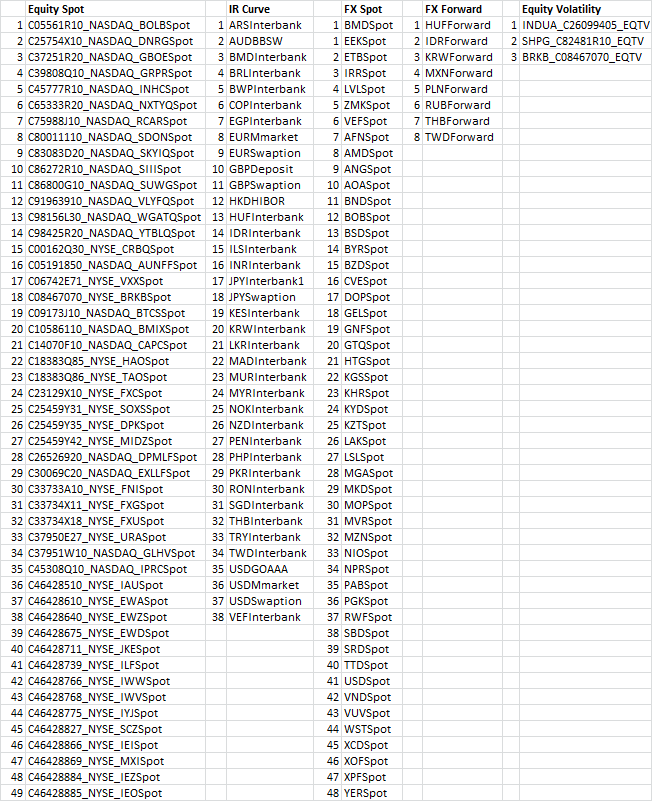
SP500\_C00000117\_EQTV,93

SP500\_C00000117\_EQTV,184

SP500\_C00000117\_EQTV,360

# Appendix G: Risk factors with no or poor quality (stale) data

Table G1: Risk factors with no or poor quality (stale) data by factor type for which proxy is not specified



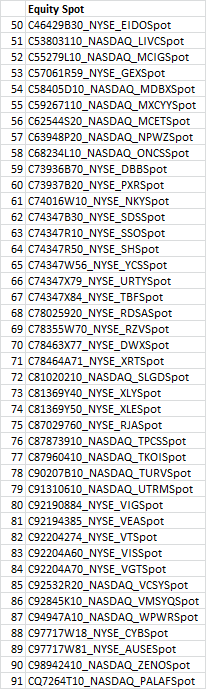


Table G2: The risk factors with gaps in historical data.

