OFR WORKING PAPER
OFFICE OF FINANCIAL RESEARCH

15-19 | October 1, 2015

Measuring the Unmeasurable: An Application of Uncertainty Quantification to Financial Portfolios

Jingnan Chen
Singapore University of Technology and Design
jingnan_chen@sutd.edu.sg

Mark D. Flood
Office of Financial Research
Mark.Flood@treasury.gov

Richard B. Sowers
University of Illinois at Urbana-Champaign
r-sowers@illinois.edu

Measuring the Unmeasurable
An application of uncertainty quantification to financial portfolios

Jingnan Chen* Mark D. Flood† Richard B. Sowers*,‡

October 1, 2015

Abstract

We extract from the yield curve a new measure of fundamental economic uncertainty, based on *McDiarmid's distance* and related methods for *optimal uncertainty quantification* (OUQ). OUQ seeks analytical bounds on a system's behavior, even where the underlying data-generating process and system response function are incompletely specified. We use OUQ to stress test a simple fixed-income portfolio, certifying its safety—i.e., that potential losses will be "small" in an appropriate sense. The results give explicit tradeoffs between: scenario count, maximum loss, test horizon, and confidence level. Unfortunately, uncertainty peaks in late 2008, weakening certification assurances just when they are needed most.

This work was initiated while Rich Sowers was a visiting Research Principal with the Office of Financial Research (OFR), Washington, DC. The authors thank Paul Glasserman, Greg Feldberg, Greg Duffee, and conference participants at the 2015 meeting of the Financial Engineering and Banking Society (FEBS) at Audencia Nantes University and the 2015 IMS-FIPS Workshop at Rutgers University for numerous helpful comments. Research support from the OFR is gratefully acknowledged. Some of this work was completed during the "Broad Perspectives and New Directions Financial Mathematics" program at the Institute for Pure and Applied Mathematics (IPAM) at the University of California at Los Angeles; the authors would like to thank IPAM for its hospitality. Any remaining errors or omissions are the responsibility of the authors alone. Comments and suggestions are welcome and should be directed to the authors. Richard Sowers was supported during part of the preparation of this work by NSF DMS 1312071.

*Singapore University of Technology and Design, Singapore
†Office of Financial Research, U.S. Department of the Treasury, Washington, D.C.
‡Department of Mathematics, University of Illinois at Urbana–Champaign, Urbana, IL

Contents

1 Introduction

This paper extracts a new measure of fundamental economic uncertainty, based on *McDiarmid's distance*, from the Treasury yield curve. McDiarmid's distance is a centerpiece of a set of methods for *optimal uncertainty quantification* (OUQ) recently developed by the engineering community. The OUQ approach seeks analytical bounds on a system's behavior, even where the underlying data-generating process and system response function are not fully specified. We adapt the methods to the problem of stress testing financial portfolios.

Uncertainty plays an important role in economics and finance, where the term traditionally refers to the statistically unmeasurable situation of Knightian uncertainty, where the event space is known but probabilities are not [18]. Full ignorance of the underlying probability structure is often an unrealistically extreme assumption, however. Uncertainty quantification originated in the engineering domain, which requires precise assurances about the probability of system failures, such as an airplane crash or bridge collapse, despite partial ignorance or ambiguity about the data-generating process. In engineered systems, failure events for the individual component parts can be rendered statistically independent by *physically* segregating the components. The probability of overall system failure then becomes an intricate, but straightforward, structured calculation of Bernoulli probabilities of subsystem and component failures.[1]

OUQ, described in Section 2.2, generalizes the approach to a broad class of problems in this middle ground, where the full probability law remains unknown, but can nonetheless be constrained in important ways.[2] OUQ requires that the system have a bounded response to each element in a collection of exogenous forcing factors. For example, in the context of a bond portfolio, McDiarmid's distance reveals the possible variation of portfolio losses (system response) resulting from a given set of interest-rate shock scenarios (input impulses). In addition, OUQ requires that the exogenous forcing factors be statistically independent. With these assumptions, the McDiarmid mathematics of concentration of measure optimally aggregate the individual bounds to produce a quantitative upper bound on the overall likelihood of extreme outcomes. Intuitively, the bounds define univariate worst-case scenarios for the response function, while the independence assumption implies that it is improbable that multiple factors will achieve their worst case bounds simultaneously.

As a stress-testing exercise, OUQ techniques allow one to "certify" portfolio soundness. More specifically, certification means asserting, with a specific confidence level, that a particular real-valued response function—the quantity of interest—will not exceed a predefined safety threshold. For example, a portfolio manager might wish to assert to clients that a drawdown during the next 12 months exceeding 25 percent of the portfolio's current market value will occur with at most a 0.1 percent probability. OUQ is a toolkit for supporting such certifications.

As described in Section 2, uncertainty played a key role in the recent rise to prominence of financial stress testing, especially as a supervisory exercise. Nonetheless, stress testing still lacks a rigorous theoretical foundation. Such an overarching framework could help identify shortcomings in current practice and help guide new policy and implementation choices. Financial stress testing instead has evolved as a practical technique for addressing the important problem of rare but severe events that can overwhelm individual financial firms and entire financial systems. This is inherently a modeling exercise, because it involves assessing the behavior of the system when extrapolated to the unusual counterfactual conditions defined by the stress scenario(s). Done carefully, stress testing can also act as a technical audit, revealing shortcomings in the design and implementation of risk models. It can also deliver an attribution analysis identifying the segments of a large portfolio most vulnerable to particular stressors.

We illustrate the use of OUQ in the case of a simple but realistic portfolio of Treasury bonds. The example demonstrates the basic feasibility of uncertainty quantification in a financial context. An important qualifier is the tightness of the bounds one can assert regarding portfolio response. Although these bounds

[1] For an example of the procedure, see the fault tree analysis in the ARP4761 airline safety standard [48, appendix D]. The maximum acceptable probability [44] of a catastrophic event per flight hour in civil aviation is 10^{-9}.

[2] For a simple example of a constraint, one might know the mean of a distribution from the law of large numbers, but not its variance or any other moments. Owhadi et al. [44] work through an illustration in which knowledge of the mean constrains the set of admissible scenarios in a certification exercise.

may seem quite loose (100 percent loss of principal), we show that OUQ techniques generate significant gains in certification confidence over the more familiar approach of Chebychev's inequality, which uses an assumption of finite variance to bound uncertainty.

For good or ill, economic uncertainty varies over time. By our measure, uncertainty tends to peak during late 2008, just when certification was needed most. In particular, our calculated McDiarmid's distance achieves its maximum at the end of 2008, when the financial crisis was most urgent. This correlation between uncertainty and crisis is unsurprising, but nonetheless symptomatic of a fundamental tension for financial stress testing, because increases in uncertainty reduce the strength of possible certifications.

2 Quantifying uncertainty

This section begins with a discussion of the case of financial stress testing, where certification has played a prominent role. Section 2.2 then describes the OUQ approach in greater detail.

2.1 Stress testing as certification

In the depths of the recent financial crisis, investors' uncertainty about asset quality drove surviving banks' stock-market valuations unrelentingly lower. Regulators responded with the emergency Supervisory Capital Assessment Program (SCAP) to staunch the crisis of confidence driven by this fundamental uncertainty.[3] In the words of Federal Reserve Chairman Bernanke [6],

> "In retrospect, the SCAP stands out for me as one of the critical turning points in the financial crisis. It provided anxious investors with something they craved: credible information about prospective losses at banks. Supervisors' public disclosure of the stress test results helped restore confidence in the banking system and enabled its successful recapitalization."

In short, the Federal Reserve sought to certify the "safety" of large bank holding companies under stressful conditions.[4] Although it was unprecedented in many ways, the SCAP largely succeeded, and it laid the foundation for the evolving program of stress testing as a centerpiece of banking supervision.

The perceived success of the SCAP during the financial crisis, renewed an emphasis on financial stress testing, which now has a central role in official supervision and industry practice. In banking, the Comprehensive Capital Analysis and Review (CCAR) and Dodd-Frank Act Stress Testing (DFAST) processes have superseded the 2009 SCAP in the United States, while the European Banking Authority has similarly instituted annual European Union-wide stress testing. These regulatory stress tests have wide-ranging implications for capital requirements and other supervisory policies. The European Insurance and Occupational Pensions Authority is currently implementing the new Solvency II stress testing process for insurance companies. For many years, the International Monetary Fund has applied its Financial Sector Assessment Programs (FSAPs) to a broad gamut of financial activities in other countries.[5]

In finance, the general class of certification problems is familiar from the widespread use of value-at-risk (VaR) in market risk analyses by practitioners and regulators; [5], [35]. A VaR analysis ordinarily starts by fixing a confidence level—the VaR α—and then works backward to estimate a safety threshold that just satisfies the confidence level. High-frequency, low-dimensionality data characterize much of modern

[3] Banking supervisors announced their intention to perform the SCAP stress test in February 2009 [11]. They published the SCAP results in May 2009 [9], which included calls for new capital injections for 10 of the 19 bank holding companies tested. See also Hirtle et al. [32] and Tarullo [49].

[4] A contemporaneous proposed statement of policy [10, p. 2] similarly noted that, "... these stress tests provided valuable information to market participants, reduced uncertainty about the financial condition of the participating bank holding companies under a scenario that was more adverse than that which was anticipated to occur at the time, and had an overall stabilizing effect."

[5] For an overview of the current state of supervisory stress testing, see Foglia [25] and Bookstaber et al. [12]. The latter also offer suggestions for a research agenda. Blaschke et al. [7] describe the FSAP methodology. Pattison [45] and Flood et al. [24] discuss the increasing reliance on data-driven methods generally in financial supervision.

risk management, creating a situation that invites such traditional statistical approaches. Even in efficient and liquid markets, however, simple Gaussian (or other stable, finite-variance) data-generating processes are typically inadequate to describe the data fully. The most fundamental difference between the VaR and OUQ approaches, however, is that VaR implementations require the assertion of a known statistical distribution as the driver of risk. OUQ similarly requires there be some reliable probability law that governs the data, but does not require that we know what that distribution is. Nonetheless, under certain general restrictions, we are able to derive precise confidence bounds on the likelihood of certain events, such as an asset drawdown in excess of available capital.

Stress testing contrasts with VaR, which collapses hazard (the contingent scenario), risk (probability of event occurrence), and impact (resulting change in a portfolio's value or other quantity of interest) into a single reduced-form risk measure. Stress testing, on the other hand, disentangles hazard, risk, and impact by removing risk from the problem and focusing directly on the possible impact of a scenario occurrence. Nonetheless, there remains an inherent tension between the severity of shocks and their likelihood of happening. Outlandish scenarios will face opposition. This trade-off is especially likely to operate when there are few scenarios and real resources are at stake (for example, in capital requirements). Supervisors acknowledged this problem early in the history of regulatory stress testing. For example, the 1996 Market Risk Amendment to the original Basel Accord notes "banks' stress tests should be both of a quantitative and qualitative nature, incorporating both market risk and liquidity aspects of market disturbances. Quantitative criteria should identify plausible stress scenarios to which banks could be exposed" [4, p. 46]. Similarly, by 2001, the International Monetary Fund had established formal process for their FSAPs, relying on "exceptional, but plausible" events [7].

We formalize the notion of financial stress testing as an exercise in certification and explore the implications of this perspective. Certification is the problem of asserting that a particular system response function will not, with probability at least $1 - \varepsilon$, exceed a particular safety threshold, α:

$$\mathbb{P}\left\{G(X) \geq \alpha\right\} \leq \varepsilon.$$

The function G represents a system response (or "quantity of interest" [42]), the event $G(X) \geq \alpha$ denotes a particular form of system failure, such as a bank failure or a portfolio credit loss exceeding a predefined risk tolerance, and $\mathbb{P}$ is a probability measure governing the data-generating process for the vector of input variables X. The practical challenge is that neither G nor $\mathbb{P}$ is usually fully known in advance.[6] The vulnerability profile of a portfolio, defined by G, can evolve quickly, especially in institutions with a large trading book. Moreover, the process emitting economic hazards, captured by $\mathbb{P}$, is typically punctuated by extreme events in a high-dimensional space. A stress event might be a terror attack, a currency crisis, or a default by a large counterparty.

Lacking a calculus to address directly the twin challenges posed by imperfect knowledge of G and $\mathbb{P}$, regulators and industry risk managers have resorted to two practical expedients. First, the response function, G, is emulated by an elaborate modeling process involving a mixture of vendor-supplied valuation and risk analytics software packages and bespoke calculation engines at the tested firms and their regulators. Second, the theoretically measurable data-generating process, $\mathbb{P}$, is replaced by a set of hand-picked scenarios, $x_1, ..., x_n$. The implementation burden here is large, and the scenario budget correspondingly small. For the CCAR exercises, the number of scenarios is $n = 2$. We do not address the practical concerns of emulation analytics and scenario selection, but explore instead the possibility of formal uncertainty quantification, applying techniques from the engineering literature.

2.2 Optimal uncertainty quantification (OUQ)

Uncertainty quantification addresses the important practical problems that lie between traditional statistical analyses in which the risks are presumed to have measurable likelihoods, and the probability-free world of

[6] Glasserman and Xu [26] adopt an alternative approach to the imperfect knowledge of G and $\mathbb{P}$, treating this as a problem of model risk and robust control [31] and applying techniques of relative entropy to measure and bound possible model errors.

pure scenario analysis. Actuarial data on system failures are often unavailable, regardless of the industry.[7] Nonetheless, probabilistic assessments of failure are a key input to important decisions, such as whether to fly the plane or close the bank. Uncertainty quantification leverages available information to derive probabilistic constraints on a quantity of interest that summarizes some key aspect of the system; [42].

Two recent papers in this literature ([39] and [44]) develop a general framework for uncertainty quantification that focuses on the assumptions and information sets that underlie such constraints. A key goal is to fully exploit available information, while avoiding inappropriate assumptions. A similar and more familiar result is Chebychev's inequality (16), which requires only that the data-generating process have finite mean and variance. In the example derived in Section 3, we demonstrate the performance improvement provided by OUQ in comparison to Chebychev.

The OUQ framework makes central use of the probability inequalities of Hoeffding [33] and related results on concentration of measure first explored by McDiarmid [40, 41].[8] McDiarmid's inequality (see equation (19)) requires two important constraints on the problem:

- The (perhaps unknown) inputs, X, to the certification problem in equation (1) must be statistically independent.
- The component-wise oscillations of the output function, $G(X)$, must have a finite diameter. That is, the largest absolute deviation in $G(X)$ for any isolated change in a single element of X must have a finite upper bound.

These requirements impose non-trivial restrictions in many applications, but they are much less restrictive than most alternative approaches. The basic intuition for the concentration-of-measure approach is that it is difficult for *independent* inputs to "conspire" to produce a worst-case outcome that would require them all to achieve their extreme (bad) values simultaneously.

OUQ traditionally focuses on mechanical rather than social systems. The presence of humans in the loop can sharply limit the applicability of OUQ, especially if the human actors have ulterior or adversarial motives and the capacity to act strategically [42]. We return to this limitation in the conclusions below. Although strategic actors play key roles in most financial institutions and markets, computer models and networks also represent important, large, and growing subsystems within the overall financial system. Examples include quantitative hedge fund strategies, algorithmic and high-frequency trading, clearinghouses and payment networks, and custom risk and pricing analytics. When isolated, these mechanical components are typically well suited to OUQ approaches.

To explore the possible benefits of the OUQ framework for financial stress testing, we apply it to the special case of a fixed-income portfolio of government bonds, where the source of risk, X, is taken to be fluctuations in the Treasury yield curve. The goal is to demonstrate the applicability of an OUQ approach to certification in a financial context and to illustrate its potential benefits in terms of uncertainty reduction. To our knowledge, this is the first application of the OUQ framework to a financial problem. Yield curve dynamics lack the high-dimensional challenges that confront a full-blown stress test of a large, complex financial institution. Indeed, we are able to use standard principal-component techniques to reduce the problem to a two-dimensional response function. Nonetheless, the example is sufficiently realistic for an initial appraisal of OUQ for financial stress testing.

3 Quantifying uncertainty for the Treasury yield curve

3.1 Framing the problem

Despite the potential for stress testing to add more formal structure to the analysis of financial vulnerabilities, practical implementations require a variety of non-trivial assumptions and choices, each of which should be

[7] The Operational Riskdata eXchange (ORX) [43] for financial institutions is an exception, accumulating data on common operational events.

[8] For additional background on concentration inequalities, see Ledoux [36] and Boucheron et al. [13].

carefully grounded. Financial stress testing practice has advanced quickly, evolving largely through ad-hoc supervisory and practitioner expedients. We argue that financial stress testing will benefit from an overarching theoretical framework to help guide implementation choices more deliberately. The theoretical foundations of stress testing are still being laid, and we hope to frame an understanding of these questions by working through the application of OUQ certification techniques in a financial context.

As discussed in Section 2, one important goal of supervisory stress tests is to ensure key financial institutions suffer manageable losses by testing whether those firms have sufficient capital to absorb even extraordinary negative shocks. In other words, supervisors would like to certify that an institution is "safe," based on a scenario analysis of balance sheet responses to a range of shocks. We formalize the problem by first assuming that there is a D-dimensional source, X, of risk factors that drive changes in the value of a portfolio, such as a loan book or a full bank balance sheet. The random variable X has an *unknown* distribution μ. There is a loss function $L : \mathbb{R}^D \to \mathbb{R}$ representing the portfolio's response to an impulse from the risk source, and we want to compute:

$$\ell \stackrel{\text{def}}{=} \mathbb{E}[L(X)] \stackrel{\text{def}}{=} \int_{x \in \mathbb{R}} L(x)\mu(dx). \tag{1}$$

To test the portfolio, a risk officer or regulator submits scenarios $\{\hat{X}_1, \hat{X}_2 \dots \hat{X}_N\}$ to the financial institution, and the institution computes $\{L(\hat{X}_1), L(\hat{X}_2) \dots L(\hat{X}_N)\}$ as well as the empirical average loss:[9]

$$\ell_N^{\mathrm{e}} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{n=1}^{N} L(\hat{X}_n). \tag{2}$$

The portfolio manager and the stress tester agree on the structural assumptions of D sources of risk and on the loss function L.

One challenge in stress testing applications is controlling the size of N.This is important in practice, because scenario analysis is typically costly, and regime shifts or data availability may limit the applicability of the model that estimates $L(\hat{X}_i)$ to a certain time interval; there may be only a finite amount of relevant historical data.[10] A strength of the OUQ framework [44] is that it allows for a finite number of scenarios. Alternately, scenario analysis may be cursed by high dimensionality, making it difficult to construct internally consistent scenarios covering all dimensions simultaneously; [23]. When stress testing a full financial institution, as in the DFAST and CCAR processes, large operational costs of analyzing each scenario effectively impose a scenario budget. In all these case, techniques that make optimal use of a limited number of scenarios should achieve a more reliable final analysis.

3.2 A model of the yield curve

To further simplify the example, we decompose the yield curve dynamics into a combination of latent factors. The empirical dynamics of Treasury yields have been heavily studied elsewhere (e.g., [14], [16], [17], [22], [30], [37], [21], [19], [46]), allowing for a relatively clear delineation of the information sets—the knowns and unknowns—supporting the analysis. The standard decomposition uses principal components analysis (PCA) to extract three sources of randomness, typically labeled as the short-term interest rate (short rate), the long rate, and the curvature of the term structure [38]. Alexander [1] provides a detailed example of the PCA procedure. This decomposition identifies the statistical factors present in the *cross section* of yields at various maturities. Duffee [20] establishes, theoretically and empirically, that the data-generating process can also embed "hidden" factors that help drive the intertemporal *dynamics* of the term structure, but that are not detectable in the cross-section alone. We work here exclusively with the factors revealed by

[9] One might also compute a number of alternative statistics, such as the expected maximum loss over a horizon or the probability that the loss exceeds a certain level; the calculations would be similar.

[10] Glasserman et al. [27, Sections 7 and 8] discuss techniques for optimal use of conditioning information to identify internally coherent stress scenarios in the multidimensional tail.

the cross-section of yields under an assumption that the data-generating process remains stable over short intervals.[11]

It is commonplace for the standard cross-sectional decomposition to explain the vast majority of the variation in rates along the yield curve.[12] There are other important approaches to dimensionality reduction and modeling of yield curve dynamics, including affine, no-arbitrage models focused on pricing applications [22] [46], and the dynamic Nelson-Siegel model [16] (Diebold and Rudebusch [17] provide an excellent overview of this literature). We opt for the traditional PCA decomposition, because the orthogonality structure it imposes on the factors provides the foundation for the extraction of fully independent factors—i.e., not merely uncorrelated—described next.

The goal at this stage is to isolate a set of risk factors, X_t, that are low dimensional (for tractability) and statistically independent (a requirement of the OUQ mathematics), while still capturing almost all the variation in the yield curve. This will allow us, in the portfolio analysis in Section 4 below, to treat the loss function as responding directly to a small set of impulses, X_t, satisfying the independence requirements of the OUQ approach. We begin by applying PCA to extract orthogonal (not yet independent) factors that approximate the full yield curve. Defining the change in value—i.e., profit or loss—for a single bond, with annual maturity $m \in \{1, 2 \ldots 30\}$ and a par value of \$1 as:

$$\xi_m(t) = \{e^{(y_t(m) - y_{t+1}(m))m} - 1\}, \tag{3}$$

The statistical behavior of the yield curve thus follows the $\mathbb{R}^M$-valued stochastic process $\{\boldsymbol{\xi}(t)\}_{t=1}^{T}$ where the vector of individual bond value fluctuations at time t is given by:

$$\boldsymbol{\xi}(t) = \begin{pmatrix} \xi_1(t) \\ \xi_2(t) \\ \vdots \\ \xi_{30}(t) \end{pmatrix} \tag{4}$$

Let Σ_e be the empirical covariance matrix of $\boldsymbol{\xi}$. We build a PCA projection operator that isolates the dominant fluctuations of $\boldsymbol{\xi}$. Let $\mathbf{E}$ be an orthonormal matrix that we can rotate so its columns, $\mathbf{e}_n$, are the eigenvectors of Σ_e. Then:

$$\mathbf{E}^T \Sigma_e \mathbf{E} = \mathrm{diag}(\lambda_1, \lambda_2 \ldots \lambda_{30}) \tag{5}$$

where $\lambda_1 \geq \lambda_2 \ldots \lambda_{30} \geq 0$. For our limited sample, the first three principal components are indeed adequate:

$$\frac{\sum_{j=1}^{3} \lambda_j}{\sum_{j=1}^{30} \lambda_j} = 99.9977;$$

As expected, the overwhelming majority of the cross-sectional covariance of the yield curve is captured by fluctuations in the direction of the first three eigenvectors of Σ_e. Figure 1 plots the first three eigenvectors of the covariance matrix. Note that the eigenvectors here are based on the covariance of prices over time, not yields. This is a departure from typical practice in the literature. For purposes of stress testing, we are most interested not in fluctuations in the yield curve itself, but rather in the dollar value of our portfolio.

More specifically, write the eigenvectors of Σ_e as the columns of $\mathbf{E}$:

$$\mathbf{E} = \begin{bmatrix} \mathbf{e}_1 \mid \mathbf{e}_2 \mid \ldots \mid \mathbf{e}_{30} \end{bmatrix}$$

[11] A comparison of our OUQ-based uncertainty measure with Duffee's [20] hidden factor(s) is an interesting question for future research.

[12] For example, Alexander [1, p. 56], shows that the first three principal components from the correlation matrix of daily spot rates for U.K. government debt account for over 99 percent of total variation. Intuitively, there is a great deal of common variation across the various tenors in the term structure, because bonds of nearby maturities are close economic substitutes whose comovement is encouraged by arbitrage.

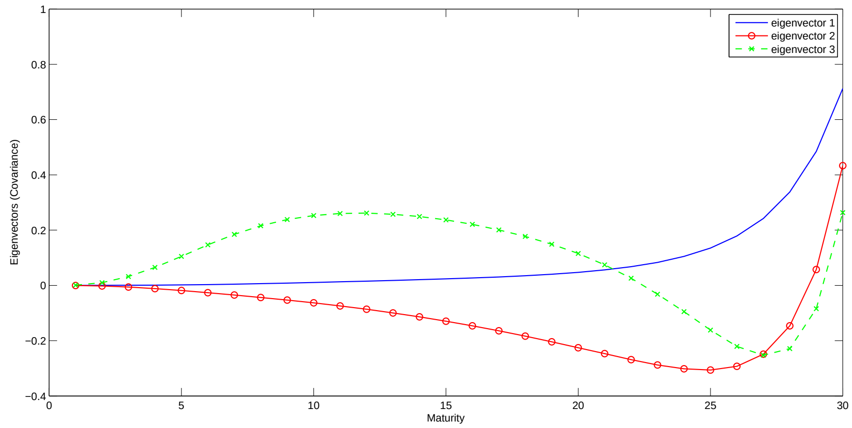

Figure 1: First Three Eigenvectors of the Covariance Matrix

so that the $\mathbf{e}_n$ sequence forms an orthornormal set in $\mathbb{R}^{30}$ and such that $\Sigma_e \mathbf{e}_m = \lambda_m \mathbf{e}_m$. This implies that we can approximate the vector of changes in bond values, $\boldsymbol{\xi}(t)$, as the combination of the empirical level and variation around it:

$$\boldsymbol{\xi}(t) \approx \tilde{\boldsymbol{\xi}}(t) \stackrel{\text{def}}{=} \bar{\boldsymbol{\xi}}_e + \sum_{j=1}^{3} \mathbf{e}_j X_t^{(j)} \tag{6}$$

where the vector $\bar{\boldsymbol{\xi}}_e$ is the empirical average of $\boldsymbol{\xi}$ (i.e., each component, $\boldsymbol{\xi}_m$, is the empirical average of the price of the bonds with maturity m), and where the j^{th} risk factor, $X_t^{(j)}$, projects onto the vector of bond value fluctuations via the j^{th} eigenvector:[13]

$$X_t^{(j)} \stackrel{\text{def}}{=} \mathbf{e}_j^T \boldsymbol{\xi}(t) \tag{7}$$

This formulation captures an optimal three-dimensional approximation of $\boldsymbol{\xi}$. The empirical covariance matrix of $(X_t^{(1)}, X_t^{(2)}, X_t^{(3)})$ is the matrix

$$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}; \tag{8}$$

This implies that the $X^{(j)}$ factors are uncorrelated. In Section 3.3, we further reduce this system to two dimensions, in preparation for the portfolio analysis in Section 4.

Collecting results: (a) there is a fixed and known matrix Σ_e that leads to the $\mathbf{e}_j$ vectors; and (b) there is a vector-valued time series X, where $\{X_t\}_{t \in \mathbb{N}}$ is an i.i.d. (over time) collection of $\mathbb{R}^3$-valued random variables with multivariate distribution (i.e., common probability law), μ. While the components of X are *orthogonal* by construction, μ does not necessarily describe three *independent* $\mathbb{R}$-valued random variables. For example, there might be tail dependencies or other conditional relationships between the factors that a simple linear correlation measure does not capture.

3.3 Deriving independent risk factors

Our main insight is to use historical data to get a better understanding of the probability law μ governing the risk factors, $X^{(j)}$. Through the principal components derivation, the covariance structure (8) implies that the $X^{(j)}$s are uncorrelated, but this is inadequate; we require statistical independence. That is, we want

[13] Because $\mathbf{E}$ is orthonormal, its inverse and transpose are equal: $\mathbf{E}^{-1} = \mathbf{E}^T$.

to write the $X^{(j)}$s as (nonlinear) functions of independently distributed random variables. The full nature of these nonlinear functions is unimportant for our purposes; we only need the bounds on the response.

We start by reducing the problem from three to two risk dimensions, namely the first two eigenvectors from the PCA derivation (5). The motivation for this is largely aesthetic, to avoid an explosion of notation in iterating the derivation to cover the third risk factor, as this would distract attention from the basic intuition for the process. However, the dimensionality reduction is further justified in the special case of our data by the fact that the remaining two principal components still account for the vast majority of the variation in the sample. We calculate that

$$\frac{\sum_{j=1}^{2}\lambda_j}{\sum_{j=1}^{30}\lambda_j} = 99.9733.$$

That is, the first two principal components explain more than 99.9 percent of the total variance. This, of course, will not be true in general.[14]

The next step is to derive independent risk factors from the uncorrelated factors, $\{X^{(1)}, X^{(2)}\}$. Assume that $\{U_1, U_2\}$ are two independent uniform $(0,1)$ random variables. Let F_1 be the first marginal of μ; i.e., the probability law of $X^{(1)}$. In other words, we map the first principal component to U_1 by inverting the cumulative distribution of $X^{(1)}$:

$$F_1(s) \stackrel{\text{def}}{=} \mu\left((-\infty, s] \times \mathbb{R}\right).$$

Define

$$\zeta_1 \stackrel{\text{def}}{=} F_1^{-1}(U_1),$$

where F_1^{-1} is the right-continuous inverse. Then

$$\mathbb{P}\{\zeta_1 \leq X^{(1)}\} = \mathbb{P}\{U_1 \leq F_1(X^{(1)})\} = F_1(X^{(1)})$$

so that ζ_1 has the same statistics (i.e., the law) as $X^{(1)}$. Next let $F_{2,x}$ be the conditional cumulative distribution function of $X^{(2)}$, given that $X^{(1)} = x$. That is,

$$\mathbb{P}\{X^{(2)} \leq t_2 | X^{(1)}\} = F_{2,X^{(1)}}(t_2).$$

Define now

$$\zeta_2 = F_{2,\zeta_1}^{-1}(U_2);$$

and then apply the inverse cumulative functions to establish the equivalence in probability between the orthogonal factors, $X_t^{(j)}$, and the independent factors, ζ_j:

$$\begin{aligned}
\mathbb{P}\{\zeta_1 \leq t_1,\, \zeta_2 \leq t_2\} &= \mathbb{E}\left[\mathbb{P}\{\zeta_2 \leq t_2|\zeta_1\}1_{\{\zeta_1 \leq t_1\}}\right] = \mathbb{E}\left[\mathbb{P}\{U_2 \leq F_{2,\zeta_1}(t_2)|\zeta_1\}1_{\{\zeta_1 \leq t_1\}}\right] \\
&= \mathbb{E}\left[F_{2,\zeta_1}(t_2)1_{\{\zeta_1 \leq t_1\}}\right] = \mathbb{E}\left[F_{2,X^{(1)}}(t_2)1_{\{X^{(1)} \leq t_1\}}\right] \\
&= \mathbb{E}\left[\mathbb{P}\{X^{(2)} \leq t_2|X^{(1)}\}1_{\{X^{(1)} \leq t_1\}}\right] = \mathbb{P}\{X^{(2)} \leq t_2,\, X^{(1)} \leq t_1\}
\end{aligned}$$

where 1_A is the indicator function for event A. We can (statistically) replace the PCA approximation in (6) with a restatement in terms of independent factors, as required by OUQ:

$$\tilde{P} \stackrel{\text{def}}{=} \bar{P} + c_1 F_1^{-1}(U_1) + c_2 F_{2,F_1^{-1}(U_1)}^{-1}(U_2); \tag{9}$$

essentially, we have written out a representation using Sklar's theorem of a copula for $(X^{(1)}, X^{(2)})$ (see [47]).

[14]This exercise is usually undertaken with yields, rather than bond prices. Alexander [1, example II.2.1] applies PCA to the U.K. Treasury yield curve, for which the first two principal components explain 97.99 percent of the variance, and later [2] extends the analysis to a stress-testing context. Similarly, when we apply PCA to the daily bond returns in our laddered portfolio (i.e., not the Treasury yield curve itself, but the exponential function given by (3), we find that, once again, the first two components explain more than 99.9 percent of the total variation (more precisely, 99.9390 percent).

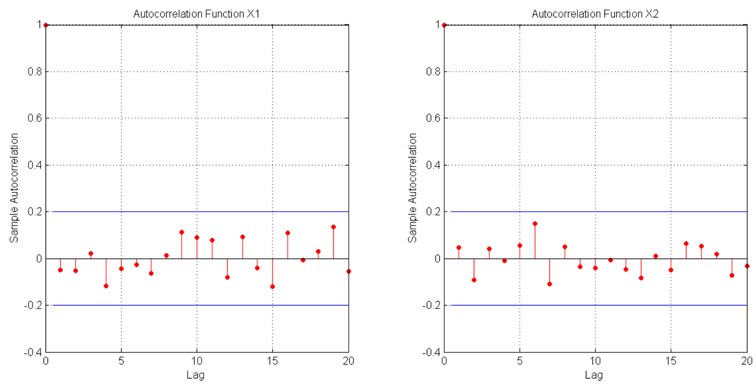

Figure 2: Autocorrelation of first and second components

Our primary goal is to introduce the OUQ methodology in the context of interest rates as a source of portfolio risk, and not to explore all the possible ramifications of this approach. Although we note, in passing, that the first two factors have autocorrelation of about 0.2 for various lags (see Figure 2), we do *not* assume *temporal* independence of the samples. Temporal dependence affects the rate at which the law of (ζ_1, ζ_2) is established from the empirical data. The possible implications of temporal (in)dependence are questions for future research. We do assume that the data-generating process remains stable within our six-month estimation windows.

3.4 Estimating the joint density

Equation (9) refactors the source of risk as a combination of two independent components. The next step is to calibrate this structure against the Treasury yield data. That is, we must empirically approximate the joint probability law of (ζ_1, ζ_2). Assume that this pair of random variables has a stable joint density $f(\zeta_1, \zeta_2)$. Given this and our data sample as a starting point, we seek the smoothest density that matches up with the empirical cumulative distribution function.

Specifically, consider the optimization problem

$$\inf \left\{ \int_{\mathbb{R}^2} |\nabla \tilde{f}(x,y)|^2 dxdy : \ \tilde{f} \geq 0, \ \int_{\mathbb{R}^2} 1_{A_{ij}}(x) \tilde{f}(x) dx = F_{X,Y}(x_i, y_j) \right\}, \tag{10}$$

where x_i and y_i are the first two principal components given by the PCA derivation and $A_{ij} = \{(x,y) : x \leq x_i, y \leq y_j\}$. Write $\tilde{f}$ for the smoothest density that fits the empirical data; in other words, $\tilde{f}$ is a density that matches the empirical observations and has the smallest gradient. This data-fitting problem is within the framework of calculus of variations (CV). We convert the CV problem into a standard quadratic program (QP) by applying a direct discretization method.

The first step in the QP approach is to choose a region D defined by boundary points $\underline{x}$, $\overline{x}$, $\underline{y}$ and $\overline{y}$ and constrain the density so that

$$\tilde{f}(x,y) = 0, \ \forall (x,y) \in D^c,$$

where $D = \{(x,y) : \underline{x} \leq x \leq \overline{x}; \underline{y} \leq y \leq \overline{y}\}$ and D^c is its complement. In other words, this truncates the infinite horizon problem to the finite region $D = [\underline{x}, \overline{x}] \times [\underline{y}, \overline{y}]$. Let X and Y represent the sets of x and y, respectively. Arrange the elements in X and Y in an increasing order. Then for each point $(x,y) \in X \times Y$, there are two associated indices i and j indicating the order of x in X and y in Y. Let I and J be the sets of indices. For $i \in I$ and $j \in J$, partition the region D as

$$h^+_{x_i} = x_{i+1} - x_i, \ h^-_{x_i} = x_i - x_{i-1}, \ h^+_{y_j} = y_{j+1} - y_j, \ h^-_{y_j} = y_j - y_{j-1},$$

where $x_i = 0$ if $i \notin I$ and $y_j = 0$ if $j \notin J$. In other words, this converts the rectangle D into an inhomogeneous grid of smaller rectangles with size $hx_i \times hy_j$.

Given the partition of D, we use the central finite difference method to approximate the gradient in the objective. Namely,

$$\tilde{f}_x(x_i, y_j) = (\tilde{f}(x_{i+1}, y_j) - \tilde{f}(x_{i-1}, y_j))/(h^+_{x_i} + h^-_{x_i}),\ \ \tilde{f}_y(x_i, y_j) = (\tilde{f}(x_i, y_{j+1}) - \tilde{f}(x_i, y_{j-1}))/(h^+_{y_i} + h^-_{y_i})$$

Hence, the objective function is a quadratic function of the decision variables $f(x_i, y_j),\ i \in I,\ j \in J$. In addition, the Hessian is a symmetric diagonally dominant matrix with positive diagonal entries, implying that it is positive semidefinite, according to Gershgorin's circle theorem. The convexity guarantees that any local optimal solution is also globally optimal. We can transform the integral constraints into a set of linear constraints by approximating an integral constraint as a sum of the products of certain decision variable average (e.g., $\frac{f(x_i,y_j)+f(x_{i+1},y_j)+f(x_i,y_{j+1})+f(x_{i+1},y_{j+1})}{4}$) and the area of the corresponding grid A_{ij} (e.g., $(x_{i+1} - x_i) \times (y_{j+1} - y_j)$). At this point, we have re-formulated the original CV problem into a convex QP problem so that any local solution is also globally optimal.

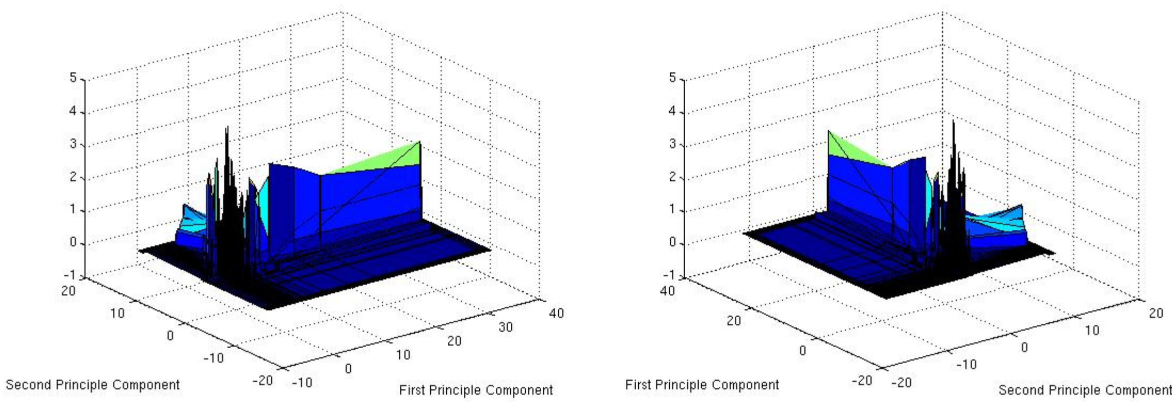

Figure 3: Joint Density Plot for the First and Second Principal Components

With this, standard optimization software will efficiently identify the optimal solution to (10), i.e., the smoothest joint density function. We use the Knitro Solver in the Tomlab platform [34] to estimate the joint density function of the two components numerically by solving the convex QP given by (10). Our calibration sample is the first two principal components of the 30-dimensional returns series for the laddered portfolio, as described in Section 3.2.[15] Figure 3 depicts the results, shown from two perspectives for clarity. As expected, the bulk of the variation occurs in the first principal component. We can now integrate the calibrated joint density function to calculate the cumulative density function and the inverse conditional cumulative function.

4 A worked example of OUQ for a fixed-income portfolio

4.1 A laddered Treasuries portfolio

We apply the methodology to the specific example of a laddered portfolio of 30 U.S. Treasury securities (bills, notes, and bonds). It is helpful to start with a relatively simple portfolio as an initial proving ground for OUQ methods in a financial context. The absence of credit risk, for example, enables an exposition of the basic techniques without some of the elaborate complications that would be necessary for a more general

[15]The computational time is 121.7 seconds to find the global optimal solution on a Scientific Linux 6.4 operating system with a duo-core 3.10GHz processor.

portfolio. The implementation of this specific example illustrates the methodology and highlights the key assumptions it requires. The example provides important lessons and guidance for a more complete exercise as a topic for future research.

The starting point for the analysis is a hypothetical laddered portfolio of 30 U.S. Treasury instruments (bills, notes, and bonds) of different annual maturities, indexed by $m \in \{1, 2 \ldots 30\}$, corresponding to constant maturities of 1 year, 2 years, and so forth. The portfolio is rebalanced each day to start with 1 unit (the total capital), invested in equal shares of \$1/30 for each maturity. For simplicity, we assume the portfolio operates on a scale sufficiently small that its impact on market prices is negligible.[16] To estimate risk-factor dynamics, we use the daily, constant-maturity Treasury yield series of Gürkaynak et al. [28], interpolated from daily, off-the-run Treasury prices [29]. We use data for $T = 100$ days in our example, from March 04, 2013 to July 24, 2013.

We normalize the calculations by assuming that each bond has par value of \$1.[17] Assuming continuous compounding, the price in dollars of the m-year bond is:

$$P_t(m) = e^{-y_t(m)m}$$

With an equally weighted investment of \$1/30, the portfolio manager can buy:

$$\frac{1/30}{P_t(m)} = \frac{1}{30} e^{y_t(m)m}$$

bonds at the start of day t, to be sold (or rebalanced) one day later at price $P_{t+1}(m)$, paying off an amount (in dollars):

$$\frac{1}{30} e^{y_t(m)m} P_{t+1}(m) = \frac{1}{30} e^{(y_t(m) - y_{t+1}(m))m}.$$

Defining the change in value for a single bond (with par value of \$1) as in Equation (3), the profit (or loss) on an individual bond m in the portfolio is $\frac{1}{30}\xi_m(t)$. Summing the effects across the portfolio, the total profit for day t is

$$P \stackrel{\text{def}}{=} \frac{1}{30} \sum_{m=1}^{30} \xi_m(t). \tag{11}$$

The statistical behavior of the portfolio—the vector of bond-by-bond profits and losses at time t—follows the same $\mathbb{R}^M$-valued stochastic process given in (4). The crux of the OUQ analysis lies in addressing what is known and unknown about this process. A basic goal is to certify portfolio safety with tighter probabilistic bounds than those available with more traditional techniques, such as Chebychev's inequality. Less powerful methodologies require more test observations to achieve the same level of assurance.

4.2 Portfolio loss

The next step is to construct a *loss* (or negative profit) function. This loss is the primary quantity of interest in the stress test, and we treat it as a response function driven by the risk factors. Clearly, the loss is the positive part of $-P$. In particular, we are interested in carefully bounding the expected loss:

$$\ell \stackrel{\text{def}}{=} \int_{x \in \mathbb{R}^2} L(x)\mu(dx).$$

[16] The assumption that the market is a fully exogenous forcing function is typical in microprudential stress testing. It is also a weakness of the current procedure. Bookstaber et al. [12] propose that stress testing should move to a "version 3.0" that explicitly considers feedback effects. Such a shift would effectively expand the scope of analysis from firm level to full financial system or subsystem.

[17] Given our standing assumption that market price fluctuations are exogenous, this is an inconsequential rescaling to simplify the notation. If, instead, the activities of the portfolio could have non-negligible impact on market prices, this rescaling assumption would require additional analysis.

where L is given by equation (13) and where $\mu \in \mathcal{P}(\mathbb{R}^2)$ is the probability law governing the risk factors, X_t, given by equation (7). We already have coarse bounds on $\tilde{P}$ via knowledge of Σ_e and Chebychev's inequality, together with loss caps at either the position level (14), or at the portfolio level (15). However, OUQ provides a tool, McDiarmid's inequality (19), that uses these loss caps to achieve even tighter control on the structure of $\tilde{P}$, and is a more reliable certification of portfolio safety.

Section 3 derived the cumulative functions for the risk factors, X_t, as two independent random variables, as in equation (9). Independence of the risk sources is one of the key requirements for application of the OUQ techniques. In addition, there is a *boundedness* requirement that the response (i.e., portfolio loss) be finite. The concentration inequalities discussed in Section 4.3 require this. In principle, regulations such as line-of-business restrictions, bans on short holdings and naked options, etc., could provide an additional bound on overall exposure, but it is ultimately the actual portfolio holdings that matter, rather than official strictures, because the latter may or may not be obeyed, particularly by an institution under stress. We emphasize that the laddered portfolio in our example meets the restriction by construction. The portfolio rebalances daily to strictly positive holdings, and the maximum daily loss cannot exceed the initial daily principal stake. This boundedness may not hold for arbitrary asset portfolios. At the firm level, shareholder losses are bounded by the limited liability of equity stakes, but *overall* loss exposures can extend to a total loss of assets—or more, for example, if the asset portfolio embeds large short positions or written put options.[18]

From equation (6) and (11), profit (or loss) is a function of the risk factors:

$$\tilde{P} \stackrel{\text{def}}{=} \bar{P} + \sum_{j=1}^{2} c_j X_t^{(j)} \tag{12}$$

where, recalling that the portfolio is rebalanced daily to hold equal \$1 stakes at each maturity, $\bar{P}$ is the average (across all bonds held) profit or loss for the portfolio:[19]

$$\bar{P} \stackrel{\text{def}}{=} \frac{1}{30} \langle \mathbf{1}, \boldsymbol{\xi}_e \rangle_{\mathbb{R}^{30}}$$

and, similarly, c_j is the average across all 30 elements of the j^{th} eigenvector:

$$c_j \stackrel{\text{def}}{=} \frac{1}{30} \langle \mathbf{1}, \mathbf{e}_j \rangle_{\mathbb{R}^{30}}$$

If K is the ceiling on losses, then from equation (12) and for $x = (x^{(1)}, x^{(2)}) \in \mathbb{R}^2$:

$$L(x) \stackrel{\text{def}}{=} \left(- \left\{ \bar{P} + \sum_{j=1}^{2} c_j x^{(j)} \right\} \right)^{+} \wedge K \stackrel{\text{def}}{=} \min \left(\max \left(-\tilde{P}, 0 \right), K \right) \tag{13}$$

The bounds should have several features. The loss of the portfolio should be between the bounds; otherwise, the position would not be established in the first place. Standard deviations are typically the first line of defense in controlling losses; portfolios are often immunized against losses within several standard deviations. Losses are typically stopped (i.e., bounded) several standard deviations away from the mean. The challenge then becomes an issue of tails: how can the loss be distributed within a large range while having controlled standard deviation? The loss might, for example, take on the extreme values near the bounds, but with very small probability. This highlights an important benefit of scenario testing—consideration of historical or synthetic scenarios in the risk factors can help explore the distribution of portfolio losses by mapping out the response function.

[18] Cherny and Madan [15] explore the implications of arbitrary portfolio holdings, and derive formal restrictions on acceptability.

[19] $\mathbf{1}$ is identically the vector of 1s in $\mathbb{R}^{30}$, so the inner product $\langle \mathbf{1}, \boldsymbol{\xi}_e \rangle_{\mathbb{R}^{30}}$ is the sum of the components of $\boldsymbol{\xi}_e$ across all maturities.

Given that a typical portfolio comprises a basket of investments, there are alternate ways to enforce bounds. For example, instead of capping losses at the level of the portfolio, one might stop losses at the level of individual holdings; that is, require that

$$X_t^{(j)} \geq -k_j \tag{14}$$

for all t and all $j \in \{1, 2\}$. In other words, impose either a hedge or risk constraint that require $X^{(j)}$ to lose no more than c_j. This implies that

$$\tilde{P} \geq \bar{P} - \sum_{j=1}^{2} c_j k_j \tag{15}$$

and

$$\max\{-P, 0\} \leq \bar{P} + \sum_{j=1}^{2} c_j k_j.$$

Note that this does not conversely imply equation (14). The OUQ estimates that we develop below depend on the rule(s) for imposing bounds on losses. More generally, the granularity of risk bounds—for example, instrument level versus portfolio level—has an effect on the certifiability of a portfolio. Although we do not pursue this direction, the calculations suggest there is an optimal way in terms of certifiability to enforce risk bounds. For a given level of uncertainty, what is the most lenient combination of instrument-level and portfolio-level bounds that will admit certification?

4.3 Concentration inequalities

We have framed the analysis of stress testing in the language of concentration inequalities, like Chebychev's familiar inequality,

$$\mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \tag{16}$$

which assert a measurable likelihood that observations will fall within a specified range. Concentration inequalities give statistical bounds on the difference between a random variable and its mean. These inequalities work by placing various bounds on the fluctuations on the random variable (for example, the hard cap on daily losses in our laddered portfolio). Concentration inequalities typically also exploit an assumption of statistical independence of stochastic inputs and can be used to give bounds on sums, as in the law of large numbers. In our example, we are able to refactor yield curve risk into a D-dimensional vector of independent components, allowing us to apply McDiarmid's inequality. This approach follows Lucas et al. [39] and Owhadi et al. [44], who use McDiarmid's result centrally in developing their OUQ technique for formal certification in an engineering context.

Based on the results in Section 4.2, we can assert a finite bound on daily portfolio losses. Based on the refactoring of risk in Sections 3.2 and 3.3, we can write the portfolio loss function in terms of two independent risk variables. These results are the chief prerequisites for the application of McDiarmid's inequality (19) to the certification problem. Formally, this is

Theorem 1 (McDiarmid). *Suppose that* $\{Z_1, Z_2 \dots Z_M\}$ *are independent random variables on some probability space* $(\Omega_\circ, \mathcal{F}_\circ, \mathbb{P}_\circ)$.

Fix a bounded function $\Phi : \mathbb{R}^M \to \mathbb{R}$ *and define* $Y \stackrel{def}{=} \Phi(Z_1, Z_2 \dots Z_M)$. *For each* $m \in \{1, 2 \dots M\}$, *define*

$$\delta_{\circ,m} \stackrel{def}{=} \sup\left\{ |\Phi(x) - \Phi(x')| : \sum_{\substack{1 \leq m' \leq M \\ m' \neq m}} |x_{m'} - x'_{m'}| = 0 \right\} \tag{17}$$

and then define

$$\Delta \stackrel{def}{=} \sqrt{\sum_{m=1}^{M} \delta_{\circ,m}^2}. \tag{18}$$

For each $\varepsilon > 0$ we then have

$$\mathbb{P}\left\{Y \geq \mathbb{E}[Y] + \varepsilon\right\} \leq \exp\left[-2\frac{\varepsilon^2}{\Delta^2}\right] \tag{19}$$

Main idea of the proof. By conditioning on an increasing number of factors, Z_m, we can define a martingale $(Y_m)_{m=0}^M$ with respect to a filtration $\{\mathcal{F}_m\}_{m=0}^M$ such that $Y_M = Y$ and $Y_0 = \mathbb{E}[Y]$ and such that

$$|Y_m - Y_{m-1}| \leq \delta_{\circ,m} \stackrel{\text{def}}{=} \sup\left\{|\Phi(x) - \Phi(x')| : \sum_{\substack{1\leq m'\leq M \\ m'\neq m}} |x_{m'} - x'_{m'}| = 0\right\}$$

Then

$$\mathbb{E}\left[e^{tY_M}\right] = \mathbb{E}\left[\mathbb{E}\left[e^{t(Y_M - Y_{M-1})}\middle|\mathcal{F}_{M-1}\right] e^{tY_{M-1}}\right] \leq \exp\left[t^2\frac{\delta^2_{\circ,M}}{8}\right]\mathbb{E}\left[e^{tY_{M-1}}\right].$$

Iterating,

$$\mathbb{E}\left[e^{tY_M}\right] \leq \exp\left[\frac{\Delta_\circ^2}{8}t^2\right]$$

where

$$\Delta_\circ \stackrel{\text{def}}{=} \sqrt{\sum_{m=1}^M \delta^2_{\circ,m}}. \tag{20}$$

We then use exponential Chebychev's inequality and optimize over t (as one does in the standard proof of large deviations upper bounds). □

The variation $\delta_{\circ,m}$ in (17) is the variation in the mth coordinate.

We wish to use McDiarmid's inequality (19) to (statistically) bound the loss incurred by a financial institution, but we have only N scenarios with which to test the loss function. Recall that ℓ^e_N is the empirical mean of the loss given by equation (2). We would like to use ℓ^e_N to bound the *population* expected loss ℓ, given by equation (1).

Our main result is

Theorem 2. *We have that*

$$\mathbb{P}\left\{\ell \geq \ell^e_N + \theta^{McDiarmid}_{N,\epsilon}\right\} \leq \epsilon$$

where

$$\theta^{McDiarmid}_{N,\epsilon} = \kappa\sqrt{\frac{1}{2N}\ln\frac{1}{\epsilon}} \tag{21}$$

with

$$\begin{aligned}\kappa^2 = \sup_{u_1,u_2,v\in\mathbb{R}} \Bigg| &\left(-\left\{\bar{P} + c_1F_1^{-1}(u_1) + c_2F^{-1}_{2,F_1^{-1}(u_1)}(u_2)\right\}\right)^+ \wedge K \\ &- \left(-\left\{\bar{P} + c_1F_1^{-1}(v) + c_2F^{-1}_{2,F_1^{-1}(v)}(u_2)\right\}\right)^+ \wedge K\Bigg| \\ + \sup_{u,v_1,v_2\in\mathbb{R}} \Bigg| &\left(-\left\{\bar{P} + c_1F_1^{-1}(u_1) + c_2F^{-1}_{2,F_1^{-1}(u_1)}(u_2)\right\}\right)^+ \wedge K \\ &- \left(-\left\{\bar{P} + c_1F_1^{-1}(u_1) + c_2F^{-1}_{2,F_1^{-1}(u_1)}(v)\right\}\right)^+ \wedge K\Bigg|;\end{aligned}$$

κ captures the relevant parts of the variation in F_1 and F_2.

To prove Theorem 2, we want to connect the setup of Section 3 to McDiarmid's inequality (Theorem 4.3). Define

$$\Phi(u_1, u_2) = \sum_{n=1}^{N} \left(-\left\{ \bar{P} + c_1 F_1^{-1}(u_1) + c_2 F^{-1}_{2, F_1^{-1}(u_1)}(u_2) \right\} \right)^+ \wedge K; \tag{22}$$

then

$$\ell^e_N \overset{(D)}{=} \Phi(U_1, U_2)$$

where $\{U_1, U_2\}$ is a collection of $2N$ iid standard uniform random variables[20]

From (20), we thus see that

$$\Delta^2_N = N \frac{\kappa^2}{N^2};$$

thus

$$\mathbb{P}\{\ell - \ell^e_N \geq L\} \leq \exp\left[-2 \frac{L^2}{\Delta^2_N} \right]. \tag{23}$$

The formula (21) identifies $\theta^{\text{McDiarmid}}_{N,\epsilon}$ as the value of L which makes the right-hand side of (23) equal to ϵ.

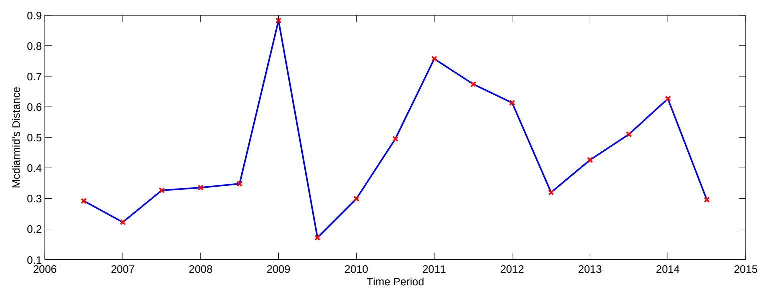

Figure 4: McDiarmid distance

In Figure 4, we plot the McDiarmid distance at 6 months (non-overlapping) windows. When the McDiarmid distance is small, the right-hand side of (23) is small, so the estimates of ℓ can be relied upon; when the McDiarmid distance is large, the right-hand side of (23) is closer to 1, so one should allow significant error in using ℓ^e_N to estimate ℓ.

By comparison, we can use a similar approach for Chebychev's inequality. Using the fact that $0 \leq L \leq K$ to conclude that L has variance of at most K, we can apply Chebychev to find that

$$\mathbb{P}\left\{ \ell \geq \ell^e_N + \theta^{\text{Chebychev}}_{N,\epsilon} \right\} \leq \epsilon$$

where

$$\theta^{\text{Chebychev}}_{N,\epsilon} = \frac{K}{\sqrt{N\epsilon}}$$

Specifically, we get the following results for the bound on the expected loss in the next quarter. The McDiarmid's distance calculated according to Theorem 4.3 is $\kappa = 0.2536$ in our example. Let $N = 64$, $K = 1$ and let ϵ vary from 0 to 1. Figure 5 plots $\theta^{\text{McDiarmid}}_{N,\epsilon}$ and $\theta^{\text{Chebychev}}_{N,\epsilon}$ as functions of ϵ, illustrating the difference between $\theta^{\text{McDiarmid}}_{N,\epsilon}$ and $\theta^{\text{Chebychev}}_{N,\epsilon}$ as ϵ varies. The McDiarmid's bound performs distinctly better than the Chebychev bound.

Finally, our choice of portfolio in this example allows for a convenient reinterpretation of the McDiarmid distance derived from equation (23) and plotted in Figure 4 as a fundamental measure of economic uncertainty

[20]The extension to higher dimensions would necessitate more complicated expressions for Φ. We leave that to the reader.

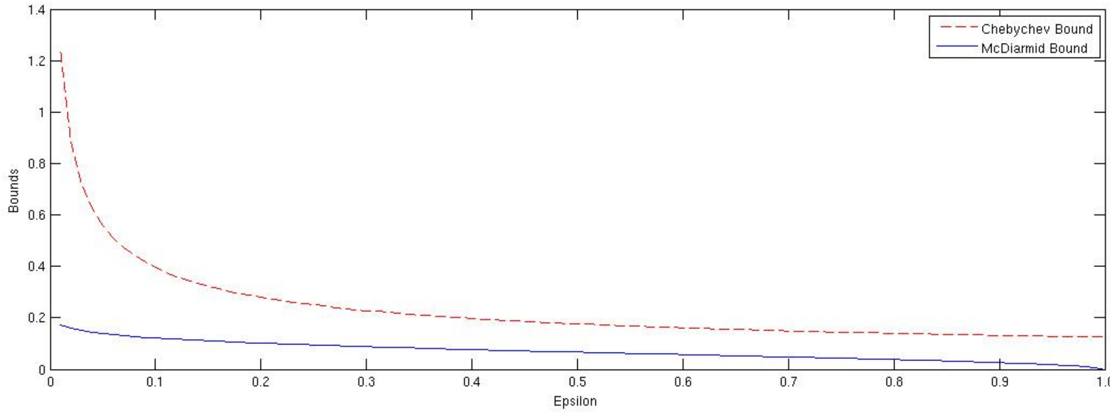

Figure 5: Bounds on the Expected Loss

embedded in the yield curve. By construction, it effectively indicates the amount of variation in the portfolio value to expect, after conditioning on the outcome of the two independent risk impulses derived in Section 3. Because the portfolio is a laddered (i.e., equally weighted) collection of Treasury securities, profits and losses on the portfolio components are a simple rescaling of the yield curve fluctuations (stated in terms of Treasury bond prices) in equation (4).

5 Conclusions

Uncertainty—the unknowns and unknowables [18]—is a central issue for financial systems, with important implications for investment decisions and economic productivity, and for the responses of system participants to stresses and surprises in markets. We apply techniques for optimal uncertainty quantification, originally developed in the engineering literature, to extract a new, fundamental measure of economic uncertainty from the prices in the Treasury yield curve. We apply this measure to the question of financial stress testing, illustrating the methodology on a simple laddered portfolio of Treasury bonds. OUQ techniques allow us to certify the probability that portfolio losses from an interest-rate shock will fall within a given ϵ bound.

An invigorated emphasis on stress testing is one of the key policy innovations to emerge from the recent financial crisis. It is now a centerpiece of banking supervision and is expanding into other areas of financial regulation. Financial firms themselves also deploy stress testing widely for their internal risk management. Despite this surge in interest, financial stress testing is still an ad hoc exercise in many respects, without a generally accepted framework to help guide technical innovations or interpret reported results. This paper contributes to the literature on stress testing by connecting it to the engineering literature on uncertainty quantification. We introduce the basic techniques of OUQ, including the key calculation of McDiarmid's distance, which provides an upper bound on the response of a system to arbitrary shocks to its inputs. In an engineering context, the McDiarmid distance forms the basis for certification guarantees that bound the probability of failure for systems and their components. We apply the measure in an economic context to extract a fundamental measure of economic uncertainty from the risk-free yield curve. We apply the measure to an example portfolio of Treasury bonds over the 2006-2014 period to demonstrate how OUQ techniques might (or might not) support certification guarantees in financial stress testing. This example allows us to ignore credit risk and focus on stress tests originating from pure market risk.

The application of McDiarmid's distance has two important prerequisites. First, the input factors influencing the system response of interest must be statistically independent of one another. In our example, the system response of interest is the economic value of the bond portfolio after the occurrence of a hypothetical shock. In engineered systems, statistical independence is often achieved from first principles by physically isolating various subsystems and their inputs. For our portfolio of Treasury securities, we approximate the separation of inputs by first performing a standard principal components decomposition of the yield curve into linearly orthogonal factors, and then further factoring their cumulative distributions to find a set of fully independent (within sample) factors. Second, the component-wise variation of the output function must have finite diameter. The absolute variation of the response to any univariate input change must have a finite upper bound; no matter what happens to an isolated input, the output response must be limited. For long bond positions, this constraint is satisfied trivially. Bond values are monotonic in discount rates, with a maximum of zero (for an infinite rate) and a minimum of undiscounted interest plus principal (for a zero rate).

Although the independence and finite-response assumptions of the McDiarmid formula appear to be satisfied for our hand-picked example, they will be less trivial in more general cases. Moreover, there are a number of other conditions that we assume to hold, either implicitly or explicitly. Some of these are relatively minor, technical assumptions that should plausibly hold for any long-only portfolio of credit-risk-free bonds. For example, we ignore day-count conventions, temporal discretization to daily observation frequencies, and divisibility to unitary (one dollar) allocations. We implicitly assume that the method of Gürkaynak et al. [28] delivers a reasonable approximation to actual bond transaction prices; those authors themselves explicitly avoid off-the-run Treasuries as excessively unpredictable. We also assume, with supporting evidence from a basic time-series analysis, that our derived risk factors are temporally independent. Although we chose to certify the portfolio's expected loss exposure at a one-quarter horizon, extending the methodology to other event horizons should be straightforward.

However, the bond portfolio example also builds in several more fundamental economic assumptions. For example, our calibration of the model uses data from the recent (2006-14) period, which includes the financial crisis. Although it is clearly useful to examine the methodology under conditions of a major stress event, it would nonetheless be useful to examine performance of the approach when applied to other asset classes, risk factors, and countries. Such extensions remain a topic for future research. A two-factor approximation to the yield curve is adequate for our Treasuries example, but it is possible that approximating other yield curves or other time periods might require additional principal components. Our method for separating orthogonal principal components into fully independent factors should be readily extensible to this higher-dimensional case, at the cost of additional layers of algebraic recursion in the derivation. We also implicitly assume that the bond portfolio is a "price taker" in the context of the stress test. The portfolio is not large enough to generate fire-sale (or other) feedback effects back to the yield curve as a consequence of a stress event; whether this assumption is justified for a large dealer is questionable.[21] Similarly, we assume the stress event does not provoke a discretionary policy response that might feed back to affect the yield curve. In general, OUQ techniques break down for systems that insert human decision-making into the process loop in significant ways, because it is impossible to impose a priori bounds on the possible extent of operator error.

The exercise supports two important conclusions. First, we can successfully apply the techniques of OUQ to the Treasury yield curve to extract McDiarmid's distance as a new, fundamental measure of macroeconomic uncertainty. The level of uncertainty is a key factor for financial stress testing, but is also relevant in many other economic and policy contexts.[22] The McDiarmid measure of uncertainty is available less frequently than certain other measures, such as that of and Baker et al. [3], but has the advantage of deriving from prices, and therefore ultimately from market participants' willingness to put capital at stake. The McDiarmid measure offers a non-redundant signal to the debate on macroeconomic uncertainty, and therefore has potential applications far beyond stress testing. This remains a topic for future research.

Second, we conclude there are daunting challenges for programmatic certification of financial stress tests. The McDiarmid distance clearly and significantly improves on the Chebychev bound as a basis for response certification. Chebychev's rule provides the most important, fully general, and distribution-free alternative to McDiarmid as a basis for bounding expected portfolio loss. Unfortunately, in the recent crisis, fundamental uncertainty, as measured by the McDiarmid's distance, spikes in late 2008, precisely when certification was most needed, as evidenced by the Fed's initiation of the Supervisory Capital Assessment Program (SCAP)

[21] Bookstaber et al. [12] argue that the absence of feedback effects is a significant limitation of the current generation of stress tests (which they refer to as "version 2.0" stress testing), especially if the goals are macroprudential.

[22] See, for example, Bloom [8] and the references therein.

process. The fact that certification would be most useful when reported for entire financial firms rather than individual investment portfolios compounds the challenges. The boundaries of the firm simultaneously represent: (1) an information firewall limiting the visibility of creditors and trading counterparties, and (2) the legal threshold at which capital regulations and bankruptcy claims will be enforced by external authorities. At this scope, the presence of human decision makers, such as managers and regulators, with significant discretionary authority to act in ways that disrupt the data-generating process defeats the basic assumptions of the OUQ methodology. Unsurprisingly, firm-level financial stress testing will probably never be a fully mechanical, automated process.

On the other hand, OUQ techniques show more promise for more narrowly isolated subsystems within the firm, such as individual portfolios or algorithms. We have demonstrated how to apply the OUQ methodology to portfolios of simple, credit-risk-free government debt, which has a relatively well understood, low-dimensional factor structure. One key to the application of the method was the exclusion from consideration of feedback effects and disruptive interventions that might shift the data-generating process. This assumption often becomes more plausible as the scope of analysis narrows to exclude human actors as well as possible interactions with other system components. It also becomes more plausible as key financial processes, such as securities trading and portfolio risk analysis, become more automated. The design and testing of these isolated algorithmic components is, or should be, an engineering exercise, to which OUQ methods could make beneficial contributions.

References

[1] C. Alexander. *Market Risk Analysis, Volume II: Practical Financial Econometrics*. John Wiley and Sons, 2008. URL http://carolalexander.org/mra_vol_2.html.

[2] C. Alexander. *Market Risk Analysis, Volume IV: Value at Risk Models*. John Wiley and Sons, 2008. URL http://carolalexander.org/mra_vol_4.html.

[3] S. R. Baker, N. Bloom, and S. J. Davis. Measuring economic policy uncertainty. Technical report, Stanford U., June 2013. URL http://www.stanford.edu/~srbaker/Papers/BakerBloomDavis_PolicyUncertainty.pdf.

[4] Basel Committee on Banking Supervision. Amendment to the capital accord to incorporate market risks. Technical Report 24, Bank for International Settlements, January 1996.

[5] Basel Committee on Banking Supervision. Messages from the academic literature on risk measurement for the trading book. Working Paper 19, BIS, January 2011. URL http://www.bis.org/publ/bcbs_wp19.pdf.

[6] B. S. Bernanke. Stress testing banks: What have we learned? Technical report, Board of Governors of the Federal Reserve, 2013. URL http://fraser.stlouisfed.org/docs/historical/bernanke/bernanke_20130408.pdf. Remarks, April 8, 2013, at "Maintaining Financial Stability: Holding a Tiger by the Tail," a Financial Markets Conference Sponsored by the Federal Reserve Bank of Atlanta.

[7] W. Blaschke, M. Jones, G. Majnoni, and S. Martinez Peria. Stress testing of financial systems: An overview of issues, methodologies, and FSAP experiences. IMF Working Paper 01/88, International Monetary Fund, June 2001.

[8] N. Bloom. The impact of uncertainty shocks. *Econometrica*, 77(3):623–685, May 2009. URL http://www.jstor.org/stable/40263840.

[9] Board of Governors of the Federal Reserve. The Supervisory Capital Assessment Program: Overview of results. Technical report, Board of Governors of the Federal Reserve, May 2009. URL http://www.federalreserve.gov/newsevents/bcreg20090507a1.pdf.

[10] Board of Governors of the Federal Reserve. Policy statement on the scenario design framework for stress testing. Technical report, Board of Governors of the Federal Reserve, November 2013. URL http://www.gpo.gov/fdsys/granule/FR-2013-11-29/2013-27009.

[11] Board of Governors of the Federal Reserve, Federal Deposit Insurance Corporation, Office of the Comptroller of the Currency, and Office of Thrift Supervision. Agencies to begin forward-looking economic assessments. Technical report, Federal Reserve, 2009. URL http://www.federalreserve.gov/newsevents/press/bcreg/20090225a.htm.

[12] R. Bookstaber, J. Cetina, G. Feldberg, M. Flood, and P. Glasserman. Stress tests to promote financial stability: Assessing progress and looking to the future. *Journal of Risk Management in Financial Institutions*, 7(1):16–25, 2014. URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2438285.

[13] S. Boucheron, G. Lugosi, and O. Bousquet. Concentration inequalities. In O. Bousquet, U. v. Luxburg, and G. Rtsch, editors, *Advanced Lectures in Machine Learning*, pages 208–240. Springer, 2004. URL http://www.econ.upf.edu/~lugosi/surveys.html.

[14] D. Brigo and F. Mercurio. *Interest Rate Models - Theory and Practice: With Smile, Inflation and Credit*. Springer, 2 edition, 2007. URL http://www.springer.com/mathematics/quantitative+finance/book/978-3-540-22149-4.

[15] A. Cherny and D. Madan. New measures for performance evaluation. *Review of Financial Studies*, 22(7):2571–2606, July 2009. ISSN 08939454. URL http://www.jstor.org/stable/40247682.

[16] F. X. Diebold and C. Li. Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130(2):337–364, February 2006. URL http://www.sciencedirect.com/science/article/pii/S0304407605000795.

[17] F. X. Diebold and G. D. Rudebusch. *Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach*. Princeton University Press, 2013. URL http://press.princeton.edu/titles/9895.html.

[18] F. X. Diebold, N. A. Doherty, and R. J. Herring, editors. *The known, the unknown, and the unknowable in financial risk management: measurement and theory advancing practice*. Princeton University Press, 2010. URL http://press.princeton.edu/titles/9223.html.

[19] G. R. Duffee. Term structure estimation without using latent factors. *Journal of Financial Economics*, 79: 507–536, 2006. URL http://www.econ2.jhu.edu/People/duffee/jfeversion.pdf.

[20] G. R. Duffee. Information in (and not in) the term structure. *Review of Financial Studies*, 24(9):2895–2934, 2011. doi: 10.1093/rfs/hhr033. URL http://rfs.oxfordjournals.org/content/24/9/2895.abstract.

[21] G. R. Duffee and R. H. Stanton. Estimation of dynamic term structure models. *Quarterly Journal of Finance*, 2(2):1250008:1–51, 2012. doi: 10.1142/S2010139212500085. URL http://www.worldscientific.com/doi/abs/10.1142/S2010139212500085.

[22] D. Duffie and R. Kan. A yield-factor model of interest rates. *Mathematical finance*, 6(4):379–406, 1996. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9965.1996.tb00123.x/abstract.

[23] M. D. Flood and G. G. Korenko. Systematic scenario selection: stress testing and the nature of uncertainty. *Quantitative Finance*, 15(1):43–59, January 2015. doi: 10.1080/14697688.2014.926018. URL http://www.tandfonline.com/doi/abs/10.1080/14697688.2014.926018.

[24] M. D. Flood, S. Kwan, and I. S. Leonova. Data for microprudential supervision of us banks. In M. S. Brose, M. D. Flood, D. Krishna, and B. Nichols, editors, *Handbook of Financial Data and Risk Information I: Principles and Context*, chapter 11, pages 426–489. Cambridge University Press, 2014. URL http://www.cambridge.org/us/academic/subjects/mathematics/mathematical-finance/handbook-financial-data-and-risk-information-i-principles-and-context-volume-1?format=HB.

[25] A. Foglia. Stress testing credit risk: a survey of authorities' approaches. *International Journal of Central Banking*, 5(3):9–45, 2009. URL http://www.ijcb.org/journal/ijcb09q3a1.pdf.

[26] P. Glasserman and X. Xu. Robust risk measurement and model risk. *Quantitative Finance*, 14(1):29–58, 2014. doi: http://dx.doi.org/10.1080/14697688.2013.822989. URL http://www.tandfonline.com/doi/abs/10.1080/14697688.2013.822989#.VTTUU2jF9CY.

[27] P. Glasserman, C. Kang, and W. Kang. Stress scenario selection by empirical likelihood. *Quantitative Finance*, 15(1):25–41, January 2015. doi: 10.1080/14697688.2014.926019. URL http://www.tandfonline.com/doi/full/10.1080/14697688.2014.926019.

[28] R. S. Gürkaynak, B. Sack, and J. H. Wright. The U.S. Treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, 54(8):2291–2304, 2007. URL http://www.federalreserve.gov/pubs/feds/2006/200628/200628abs.html.

[29] R. S. Gürkaynak, B. Sack, and J. H. Wright. The U.S. Treasury yield curve: 1961 to the present - screen reader format. Technical report, Federal Reserve Board, 2014. URL http://www.federalreserve.gov/econresdata/researchdata/feds200628_1.html. Data series (updated daily) to accompany the J. of Monetary Economics article.

[30] J. D. Hamilton and J. C. Wu. Identification and estimation of Gaussian affine term structure models. *Journal of Econometrics*, 168(2):315–331, June 2012. URL http://www.sciencedirect.com/science/article/pii/S0304407612000450.

[31] L. P. Hansen and T. J. Sargent. *Robustness*. Princeton University Press, 2007. URL http://press.princeton.edu/titles/8535.html.

[32] B. Hirtle, T. Schuermann, and K. Stiroh. Macroprudential supervision of financial institutions: Lessons from the SCAP. Staff Report 409, Federal Reserve Bank of New York, November 2009. URL www.ny.frb.org/research/staff_reports/sr409.pdf.

[33] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[34] K. Holmström, A. O. Göoran, and M. M. Edvall. User's guide for Tomlab/Knitro v6. Technical report, Tomlab Optimization, April 2009. URL `http://tomopt.com/docs/TOMLAB_KNITRO.pdf`.

[35] P. Jorion. *Value at Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill, 3rd edition, 2006. URL `http://shop.mheducation.com/mhshop/productDetails?isbn=0071464956`.

[36] M. Ledoux. *The Concentration of Measure Phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2001. URL `http://www.ams.org/mathscinet-getitem?mr=1849347`.

[37] R. B. Litterman and J. Scheinkman. Common factors affecting bond returns. *Journal of Fixed Income*, 1(1): 54–61, 1991. URL `http://www.iijournals.com/doi/abs/10.3905/jfi.1991.692347`.

[38] R. B. Litterman, J. Scheinkman, and L. Weiss. Volatility and the yield curve. *Journal of Fixed Income*, 1(1): 49–53, 1991. URL `http://www.iijournals.com/doi/abs/10.3905/jfi.1991.692346`.

[39] L. Lucas, H. Owhadi, and M. Ortiz. Rigorous verification, validation, uncertainty quantification and certification through concentration-of-measure inequalities. *Computer Methods in Applied Mechanics and Engineering*, 197(51-52):4591–4609, 2008. URL `http://www.psaap.caltech.edu/publications/files/CMAME-D-07-00486-Revision-4.pdf`.

[40] C. McDiarmid. On the method of bounded differences. *Surveys in Combinatorics (London Mathematical Society Lecture Note Series)*, 141:148–188, 1989.

[41] C. McDiarmid. Concentration. In *Probabilistic methods for algorithmic discrete mathematics*, pages 195–248. Springer, 1998.

[42] National Research Council. Assessing the reliability of complex models: Mathematical and statistical foundations of verification, validation, and uncertainty quantification. Technical report, National Academies Press, 2012. URL `http://www.nap.edu/catalog.php?record_id=13395`.

[43] Operational Riskdata eXchange Association. Operational risk reporting standards: Edition 2011. Technical report, ORX, July 2012. URL `http://www.orx.org/Pages/ORXStandards.aspx`.

[44] H. Owhadi, C. Scovel, T. J. Sullivan, M. McKerns, and M. Ortiz. Optimal uncertainty quantification. *SIAM Review*, 55(2):271–345, 2013. URL `http://epubs.siam.org/doi/abs/10.1137/10080782X?journalCode=siread`.

[45] J. C. Pattison. Data-driven regulation and financial reform: one perspective from industry on the financial crisis. In M. S. Brose, M. D. Flood, D. Krishna, and B. Nichols, editors, *Handbook of Financial Data and Risk Information I: Principles and Context*, chapter 5, pages 148–173. Cambridge University Press, 2014. URL `http://www.cambridge.org/us/academic/subjects/mathematics/mathematical-finance/handbook-financial-data-and-risk-information-i-principles-and-context-volume-1?format=HB`.

[46] M. Piazzesi. Affine term structure models. In Y. Ait-Sahalia and L. P. Hansen, editors, *Handbook of Financial Econometrics, Volume 1: Tools and Techniques*, volume 1, chapter 12, pages 691–766. Elsevier North Holland, 2010. URL `http://www.sciencedirect.com/science/book/9780444508973`.

[47] M. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231, 1959.

[48] Society of Automotive Engineers (SAE). Guidelines and methods for conducting the safety assessment process on civil airborne systems and equipment. Aerospace Recommended Practice ARP4761, SAE, December 1996. URL `http://www.ida.liu.se/~TDDD29/downloads/Ariane/ARP4761guideforSAprocess.pdf`.

[49] D. K. Tarullo. Stress testing after five years. Technical report, FRB, June 2014. URL `http://www.bis.org/review/r140627a.pdf`. Speech at the Federal Reserve Third Annual Stress Test Modeling Symposium, Boston, Massachusetts, 25 June 2014.