Scenario Analysis for Derivatives Portfolios via Dynamic Factor Models

Martin B. Haugh
Imperial College Business School
Imperial College, London
martin.b.haugh@gmail.com

Octavio Ruiz Lacedelli
Department of IE&OR
Columbia University
or2200@columbia.edu

First Draft: July 22, 2019

Abstract

A classic approach to financial risk management is the use of scenario analysis to stress test portfolios. In the case of an S&P 500 options portfolio, for example, a scenario analysis might report a P&L of −\$1m in the event the S&P 500 falls 5% and its implied volatility surface increases by 3 percentage points. But how accurate is this reported value of −\$1m? Such a number is typically computed under the (implicit) assumption that all other risk factors are set to zero. But this assumption is generally not justified as it ignores the often substantial statistical dependence among the risk factors. In particular, the expected values of the non-stressed factors conditional on the values of the stressed factors are generally non-zero. Moreover, even if the non-stressed factors were set to their conditional expected values rather than zero, the reported P&L might still be inaccurate due to convexity effects, particularly in the case of derivatives portfolios. A further weakness of this standard approach to scenario analysis is that the reported P&L numbers are generally not back-tested so their accuracy is not subjected to any statistical tests. There are many reasons for this but perhaps the main one is that scenario analysis for derivatives portfolios is typically conducted without having a probabilistic model for the underlying dynamics of the risk factors under the physical measure P. In this paper we address these weaknesses by embedding the scenario analysis within a dynamic factor model for the underlying risk factors. Such an approach typically requires multivariate state-space models that can model the real-world behavior of financial markets where risk factors are often latent, and that are sufficiently tractable so that we can compute (or simulate from) the conditional distribution of unstressed risk factors. We demonstrate how this can be done for observable as well as latent risk factors in examples drawn from options and fixed income markets. We show how the two forms of scenario analysis can lead to dramatically different results particularly in the case of portfolios that have been designed to be neutral to a subset of the risk factors.

1. Introduction

Financial risk management is a key function throughout the finance and insurance industries. At the aggregate level banks, investments firms and insurance companies all need to understand their exposure to adverse movements in the financial markets. This is also true within these firms at the level of a portfolio manager (p.m.) or trading desk where it is important to operate within certain risk constraints. One of the main approaches to financial risk management is the use of scalar risk measures such as Value-at-Risk (VaR) or Conditional Value-at-Risk (CVaR) to measure the riskiness of a given portfolio over a given time horizon such as one day or one week. While VaR (and to a lesser extent CVaR) are very popular and often mandated by regulators, this approach does have serious weaknesses. First and foremost, it can be extremely difficult to estimate the VaR of a portfolio and this is particularly true for portfolios containing complex derivative securities, structured products, asset-backed securities etc. Even when the VaR can be estimated accurately, it is impossible to adequately characterize the risk of a portfolio via a single scalar risk measure such as its VaR. In addition, a VaR does not identify the risk factors within the portfolio nor the exposure of the portfolio to those factors. One way to mitigate this for a derivatives portfolio is via the so-called Greeks such as the delta, vega and theta of an options portfolio. But the Greeks are only local risk measures and can be extremely inaccurate for large moves in the corresponding risk factors. Such moves, of course, are the principal concern in risk management.

It is no surprise then that scenario analysis is one of the most popular approaches to risk management. While there are many forms of scenario analysis, the basic idea is to compute the P&L of the portfolio under various combinations of stresses to one or more of the risk factors (or securities) driving the portfolio's valuation. Given these P&L numbers, the risk management team can assess whether or not the portfolio is too exposed to any of the risk factors and if so, what actions to take in order to reduce the exposure. In the case of an S&P 500 options portfolio, for example, a risk manager might report a P&L of -$1m in the event that the S&P 500 falls 5% and its implied volatility surface increases by 3 points.

One supposed advantage of scenario analysis, particularly for derivatives portfolios, is that a probabilistic model for the risk factor dynamics is not required. In the example above, for example, a model is not required to assess how likely is the scenario that the S&P 500 falls approx. 5% and its implied volatility surface increases by approx. 3 points. Instead the p.m. or risk management team can use their experience or intuition to assess which scenarios are more likely. For example, it is very unlikely indeed that a large drop in the S&P 500 would be accompanied by a drop in implied volatilities and so the experienced risk manager will know that such a scenario can be discounted. Nonetheless, this approach is not scientific and we are led to question just how accurate the reported value of -$1m is in the original scenario above?

In fact we argue in this paper that such a scenario P&L number can be very inaccurate. First, such a number is typically computed under the (implicit) assumption that all other risk factors, i.e. all risk factors besides the underlying and parallel shifts in the volatility surface in our example above, are set to zero. But this assumption is generally not justified as it ignores the often substantial statistical dependence among the risk factors. In particular, the expected values of the non-stressed factors conditional on the values of the stressed factors, are generally non-zero. Second, even if the non-stressed factors were set to their conditional expected values rather than zero, the reported P&L might still be inaccurate due to convexity effects in the case of derivatives portfolios whose values typically depend in a non-linear manner on the risk factors. A further weakness of this *standard approach* to scenario analysis (for derivatives portfolios) is that the reported

P&L numbers are typically not back-tested so their accuracy is not subjected to any statistical tests. There are many reasons for this but perhaps the main one is that scenario analysis for derivatives portfolios is typically conducted without having a probabilistic model for the underlying dynamics of the risk factors. A second reason is that none of the considered scenarios ever actually occur since they're zero probability events. After all, the probability of the S&P 500 falling exactly 5% and its entire implied volatility surface increasing by exactly 3 volatility points is zero so one can't immediately reject the number of -$1m.

This is in contrast to the use of VaR where it is quite standard to count the so-called VaR *exceptions* and subject them to various statistical tests that are used to determine the accuracy of the VaR estimation procedure. But the back-testing of VaR is inherently easier as it only requires the use of univariate time-series models for the portfolio P&L. In contrast, back-testing scenario analysis would require multivariate time-series models for the various risk-factors and they are considerably more complicated to estimate and work with than their univariate counterparts. Moreover risk-factor returns are often latent and therefore necessitate the use of *state-space* models. This adds a further complication to back-testing since after the fact one can only estimate (rather than know with certainty) what the realized latent risk factor returns were.

In this paper we attempt to address these weaknesses with scenario analysis by embedding it within a dynamic factor model for the underlying risk factors. Such an approach requires multivariate time series or state-space models that can model the real-world behavior of financial markets, e.g volatility clustering, and that are sufficiently tractable so that we can compute and simulate[1] from the distribution of unstressed risk factors conditional on the given scenario. We demonstrate how this can be done for observable as well as latent risk factors in examples drawn from options and fixed income markets. We also show how the two forms of scenario analysis can lead to dramatically different results particularly in the case of portfolios that have been designed to be neutral to a subset of the risk factors. The twin goals of this paper then are (i) to highlight just how inaccurate the standard approach to scenario analysis can be and (ii) to argue for a more accurate and scientific approach whereby the reported P&L numbers of a given model can be back-tested and therefore rejected if necessary. The particular models that we use in our numerical applications are intended to simply demonstrate that it is possible and important to embed scenario analysis in a dynamic factor model framework. As such they are merely a vehicle for demonstrating our approach and we certainly don't claim they are the "best" such models or that they would be difficult to improve upon. We emphasize at this point that when we refer to standard scenario analysis we have in mind scenario analysis in the context of derivatives portfolios, with perhaps the prototypical example being a portfolio of options and futures on some underlying index such as the S&P 500.

The remainder of this paper is organized as follows. We briefly discuss the literature related to scenario analysis in Section 2 and then in Section 3 we introduce standard scenario analysis and discuss in further detail its many weaknesses, again with derivatives portfolios in mind. We show how scenario analysis can be embedded in a dynamic factor model framework in Section 4 and in Section 5 we discuss how this framework can be used to evaluate the performance of standard scenario analysis. We then consider an application to a portfolio of options on the S&P 500 in Section 6. In Section 7 we discuss statistical approaches for validating a dynamic factor model in the context of scenario analysis and we conclude in Section 8 where we also outline some directions for future research. Some technical details and additional results are relegated to

[1]One of the advantages of using simulation is that we can easily estimate other risk measures besides the expected P&L in a given scenario. For example we could estimate the P&L's standard deviation or VaR conditional on the scenario.

the appendices. Appendix D in particular presents a further application of our approach where we consider scenario analysis in the context of a non-derivatives portfolio, namely of portfolios of U.S. Treasury securities.

2. Literature Review

Following the European Central Bank [14] and Rebonato [29], it is often useful to distinguish between four different types of scenarios, namely historical, hypothetical, probabilistic, and reverse-engineered scenarios. Historical scenarios apply market moves that occurred in the past to today's portfolio. Common examples are the market moves from the times of the Lehman bankruptcy, the LTCM crisis or the bursting of the dot-com bubble. Hypothetical scenarios apply to the current portfolio the market moves that might be expected to occur in the event of some subjectively defined scenario such as the break up of the Euro. Probabilistic scenarios are constructed by fitting probabilistic models to historical data and using the estimated joint distribution to develop and analyze scenarios of interest. The work in this paper is very much based on the probabilistic approach to scenario analysis and then largely with derivative portfolios in mind. Finally, reverse-engineered scenarios are scenarios that have been identified as the most likely to produce losses at or exceeding some threshold level in a given portfolio. It is worth noting that each of the four scenario types can and often do include macro-financial changes as well as market returns. It is also worth emphasizing that the boundaries between the four scenario types are often blurred and that many approaches have a hybrid flavor. Reverse-engineered scenarios often use a probabilistic model that has been fitted to historical data, e.g. Glasserman, Kang and Kang [17]. Moreover, even the most hypothetical scenario analysis will be informed at least to some extent by statistical considerations relating to the frequency of similar or related events in the past.

Regardless of the particular approach, the Basel Committee on Banking Supervision [25] recommend that scenarios be *plausible* and *severe* as well as *suggestive* of risk-reducing actions. Breuer et al. [7] find useful stress scenarios by defining a region of plausibility for the distribution of the risk-factors and systematically look for the scenario that results in the worst portfolio loss over this region. They also consider the *partial-scenario* problem where they only stress a subset of the risk-factors and show that to maximize plausibility we should set non-stressed systematic risk factors to their conditional (on the values of the stressed factors) means. This observation is central to the approach we take in this paper where we emphasize the inaccuracies that can result from implicitly setting non-stressed factors to zero which is often what is done in practice for derivatives portfolios. Flood and Korenko [15] present a systematic approach for scenario selection and use a plausible joint probability distribution together with a grid search over the space of the risk factors. In contrast, Golub, Greenberg and Ratcliffe [19] select scenarios using the hypothetical approach. In an interesting recent development Rebonato, e.g. [28, 29], proposes the use of graphical models to select scenarios. This is a more hybrid approach involving elements of both the probabilistic and hypothetical approaches. Other work relevant to scenario analysis includes Alfaro and Drehmann [1], Quagliarello [27] and Bonti et al. [5], while Borio et al. [6] provide a review of current practice.

The probabilistic approach to scenario analysis often requires the construction of factor models and certainly there is a rich history of factor models in finance going back to the CAPM and later the Fama-French 3-factor model, for example. While most such factor models have an equity focus, other asset classes have also been considered. For example Diebold and Li [13] construct a 3-factor model for the U.S. yield curve. Other approaches include the use of principal components analysis (PCA), e.g. Cont and Fonseca

[9] develop a PCA-based factor model for the implied volatility surfaces for equities. It is often necessary to include a time-series component to these models to model the non-iid nature of financial assets and risk factors. There are many approaches to this, some of which are more ad-hoc in nature, e.g. fitting univariate GARCH models to the principal components. More generally there is now a rich history in multivariate time series modeling in finance with a focus on multivariate extensions of GARCH models and (perhaps to a lesser extent) on state-space models which are required for modeling latent variable dynamics. This literature is too vast for us to consider here and instead we refer to leading textbooks such as [23] and [34] and the discussions and references contained therein.

3. Preliminaries and Standard Scenario Analysis

We assume we have a fixed portfolio of securities which in principle could include any combination of securities – derivatives or otherwise – from any combination of asset classes. In practice, however, we are limited to reasonably liquid securities for which historical price data is available. Moreover, because of the many difficulties associated with modelling across asset classes, we mainly have in mind portfolios that contain only securities from just one or two closely related asset classes. Examples include portfolios of options and futures on the S&P 500 or portfolios of US Treasury securities. We consider such portfolios in the numerical experiments of this paper but it should be possible to handle more complex portfolios albeit at the cost of requiring more sophisticated models. These more complex examples might include portfolios consisting of options and equity positions on US stocks, portfolios of spot and option positions on the major FX currency pairs, or even more ambitiously, portfolios consisting of CDS and CDO positions on US credits.

We assume then we are given a fixed portfolio and the goal is to perform some form of scenario analysis on this portfolio. We let V_t denote the time t value so that the portfolio P&L at time $t+1$ is $\Delta V_t := V_{t+1} - V_t$. In the financial context, we have in mind that time is measured in days so that ΔV_t would then be a daily P&L. We assume V_t is known at time t but ΔV_t is random. A fundamental goal of risk managers then is to understand the distribution of ΔV_t. This is required, for example, to estimate the VaR or CVaR of the portfolio.

As is standard in the risk management literature, we will assume the portfolio value V_t is a function of n risk factors whose time t values we denote by $\mathbf{x}_t \in \mathbb{R}^n$. It therefore follows that $V_t = v(\mathbf{x}_t)$ for some function $v : \mathbb{R}^n \to \mathbb{R}$. The components of $\mathbf{x}_t$ might include stock prices in the case of equity portfolios, yield-to-maturities for fixed income portfolios or implied volatility levels for a number of strike-maturity combinations in the case of an equity options portfolios. While $\mathbf{x}_t$ is random, we assume it is $\mathcal{F}_t$-adapted where $\mathbb{F} := \{\mathcal{F}_t\}$ denotes the filtration generated by all relevant and observable security prices and risk factors in the market. We define the change in risk factor vector $\boldsymbol{\Delta}\mathbf{x}_t := \mathbf{x}_{t+1} - \mathbf{x}_t$ so that

$$\Delta V_t(\boldsymbol{\Delta}\mathbf{x}_t) = v(\mathbf{x}_t + \boldsymbol{\Delta}\mathbf{x}_t) - v(\mathbf{x}_t) \tag{1}$$

where we have omitted the dependence of ΔV_t on $\mathbf{x}_t$ in (1) since $\mathbf{x}_t$ is known at time t and so the uncertainty in ΔV_t is driven entirely by $\boldsymbol{\Delta}\mathbf{x}_t$.

3.1. Standard Scenario Analysis

In a standard scenario analysis (SSA hereafter), the risk manager would identify various stresses to apply to $\boldsymbol{\Delta}\mathbf{x}_t$ in (1). For example, such stresses might include parallel shifts or curvature changes in the yield curve for a fixed income portfolio. In the case of a portfolio of futures and options on the S&P 500, these stresses might include shifts to the value of the underlying, i.e. the S&P 500, as well some combination of parallel shifts to the implied volatility surface and a steepening / flattening of the skew or term structure of implied volatilities.

When critiquing SSA it is convenient to work with a factor model for the risk factors $\mathbf{\Delta x}_t$. Such a factor model might take the form

$$\mathbf{\Delta x}_t = \mathbf{B}\mathbf{f}_{t+1} + \boldsymbol{\epsilon}_{t+1}, \qquad t = 0, 1, \ldots \tag{2}$$

where:

- $\mathbf{f}_{t+1} \in \mathbb{R}^m$ is the **common risk factor (c.r.f.)** random return vector. Some of these factor returns may be latent.
- $\mathbf{B} = [\mathbf{b}_1 \ldots \mathbf{b}_m] \in \mathbb{R}^{n \times m}$ is the matrix of factor loadings and $\mathbf{b}_i \in \mathbb{R}^n$ is the i^{th} column of $\mathbf{B}$.
- The $\boldsymbol{\epsilon}_{t+1}$'s $\in \mathbb{R}^n$ are an i.i.d. sequence of zero-mean random vectors representing idiosyncratic error terms that are assumed to be independent of the common factors returns.

Consider for example a portfolio of US Treasury securities. Then $\mathbf{\Delta x}_t$ would naturally represent changes in yield-to-maturities with each component of $\mathbf{\Delta x}_t$ corresponding to a different maturity. If the first common risk factor represented parallel shifts of the yield curve, we would fix $\mathbf{b}_1$ to be a vector of ones. If we then wished to consider a scenario where all yields increase by 20 basis points, we would set $f_{1,t+1}$, the first component of $\mathbf{f}_{t+1}$, equal to +20 bps and set the other $m-1$ components of $\mathbf{f}_{t+1}$ (as well as $\boldsymbol{\epsilon}_{t+1}$) to zero. The portfolio P&L would then be computed via (1) with $\mathbf{\Delta x}_t$ determined by the scenario and (2).

More generally, we can define a scenario by jointly stressing any number $k \leq m$ of the c.r.f. returns. Consider again our example of an options and futures portfolio on the S&P 500. In this case suppose the first component of $\mathbf{\Delta x}_t$ refers to the log-return on the S&P 500 between days t and $t+1$ with the other components of $\mathbf{\Delta x}_t$ then representing[2] changes in the implied volatilities (between days t and $t+1$) for the various strike-maturity option combinations that appear in the portfolio. If $f_{1,t+1}$ represents changes in the S&P 500 spot price then[3] $\mathbf{b}_1 = [1\,0 \cdots 0]^\top$. Similarly, if $f_{2,t+1}$ represents parallel shifts to the implied volatility surface then the second column of $\mathbf{B}$ would be $\mathbf{b}_2 = [0\,1 \cdots 1]^\top$. We can now consider a scenario where $f_{1,t+1}$ and $f_{2,t+1}$ are simultaneously stressed. For example, a scenario of interest might be one where $(f_{1,t+1}, f_{2,t+1}) = (-5\%, +10)$ corresponding to a 5% fall in the S&P 500 and a 10 volatility point increase across its entire implied volatility surface. Once again, under the SSA approach the portfolio P&L can be computed via (1) with $\mathbf{\Delta x}_t$ determined by (2) where $(f_{1,t+1}, f_{2,t+1}) = (-5\%, +10)$ and all other components of $\mathbf{f}_{t+1}$ and $\boldsymbol{\epsilon}_{t+1}$ set to zero.

In practice, a matrix of scenario P&L's might be computed as above and in fact multiple two- or even three-dimensional matrices can be computed corresponding to the simultaneous stressing of $k = 2$ or $k = 3$ different c.r.f. returns. It is important to emphasize that in the context of derivatives portfolios, the typical risk / portfolio manager employing SSA does not have an explicit model like (2) at hand nor does she need one. In fact, most of the modeling effort in the derivatives space goes into developing dynamic pricing models

[2]We are assuming that the main risks in the portfolio are underlying and volatility risks. If for example, the portfolio was exposed to substantial dividend or interest rate risk, which is quite possible in an S&P options portfolio, then additional risk factors for these risks should be included.

[3]We would also have $\mathrm{Var}(\boldsymbol{\epsilon}_{1,t}) = 0$ since this would be an instance where a component of $\mathbf{\Delta x}_t$ coincides with one of the c.r.f.'s.

under an equivalent martingale measure Q and then developing numerical procedures for computing prices and the so-called Greeks within these models. It seems very little effort is spent developing dynamic models for the risk factors driving these prices under the data-generating measure P. The main point of this article then is to highlight the many weaknesses of SSA for derivatives portfolios and to argue for a more systematic and scientific approach to it. We can do this by *explicitly* embedding SSA in a factor model such as (2) and computing the scenario P&L by also accounting for the dependence structure in (2) and not blindly setting $\boldsymbol{\epsilon}_{t+1}$ and the unstressed components of $\mathbf{f}_{t+1}$ to zero.

3.2. Problems with Standard Scenario Analysis

Before proceeding, we first expand on the many weaknesses of SSA. They include:

1. A factor model of the form (2) is rarely explicitly stated, particularly for portfolios of derivative securities. In fact, it may be the case that only a subset of the factors, say the first $l \leq m$, are ever considered for stressing. In that case SSA works with a "model" of the form

$$\boldsymbol{\Delta}\mathbf{x}_t = \mathbf{B}_{1:l,t}\mathbf{f}_{1:l,t+1} \tag{3}$$

   where $\mathbf{B}_{1:l,t}$ refers to the matrix containing the first l columns of $\mathbf{B}$ and $\mathbf{f}_{1:l,t+1}$ the vector containing the first l elements of $\mathbf{f}_{t+1}$. The important feature of (3) is that probability distributions are not specified and in fact play no role in it. It is therefore not a probabilistic model for the risk factor returns $\boldsymbol{\Delta}\mathbf{x}_t$.

2. Let $\mathbf{s} = (s_1, \ldots, s_k)$ where $s_1 \leq s_2 \leq \cdots \leq s_k \leq l$. We then let $\mathbf{f}_{\mathbf{s},t+1} := (f_{s_1,t+1}, \ldots, f_{s_k,t+1})$ denote the subset of c.r.f.'s that are stressed under a given scenario. SSA then implicitly assumes

$$\mathbb{E}_t[\mathbf{f}_{\mathbf{s}^c,t+1} \mid \mathbf{f}_{\mathbf{s},t+1}] = \mathbf{0} \tag{4}$$

   where $\mathbf{s}^c$ denotes the complement of $\mathbf{s}$ so that $\mathbf{f}_{\mathbf{s}^c,t+1}$ denotes the non-stressed c.r.f. returns in the scenario. (We use $\mathbb{E}_t[\cdot]$ throughout to denote expectations that are conditional on $\mathcal{F}_t$.) But (4) is typically not justified and can lead to a very inaccurate estimated P&L for the scenario.

3. Following on from the previous point, an obvious solution would be to set the unstressed factors $\mathbf{f}_{\mathbf{s}^c,t+1}$ equal to their conditional expectation $\mathbb{E}_t[\mathbf{f}_{\mathbf{s}^c,t+1} \mid \mathbf{f}_{\mathbf{s},t+1}]$ when estimating the scenario's P&L. While this should be an improvement over SSA, it ignores the uncertainty in $\boldsymbol{\epsilon}_{t+1}$ and $\mathbf{f}_{\mathbf{s}^c,t+1} \mid (\mathcal{F}_t,\, \mathbf{f}_{\mathbf{s},t+1})$. This uncertainty may be significant, particularly for portfolios containing securities whose values depend non-linearly on $\boldsymbol{\Delta}\mathbf{x}_t$. But even setting $\mathbf{f}_{\mathbf{s}^c,t+1} = \mathbb{E}_t[\mathbf{f}_{\mathbf{s}^c,t+1} \mid \mathbf{f}_{\mathbf{s},t+1}]$ is not a straightforward task, however, as it requires a model for the common risk factor return dynamics.

4. Finally, SSA does not lend itself to rigorous back-testing and so SSA is not open to statistical rejection and so in the language of Popper [26], SSA is not falsifiable. There are several reasons for this. First, each of the scenarios considered by an SSA are zero probability events and none of the considered scenarios will have actually occurred on day $t+1$. If this were the only problem, then it would be easy

to overcome. Specifically, on day $t+1$ we could "see" exactly what the return in the S&P 500 and what the parallel change in the implied volatility surface were over the period $[t, t+1]$. We could then rerun the scenario analysis for exactly this scenario, i.e. the scenario that transpired, and then compare the estimated and realised P&L's.

The problem with this, however, is that we cannot directly observe the actual parallel change in the implied volatility surface that transpired. This is because this c.r.f. would be a *latent* c.r.f. and so could only be estimated / inferred. But to do this a probabilistic model would be required and as we have noted, SSA often proceeds without a probabilistic model.

Following on from point 4 above, any probabilistic factor model as in (2) would surely be statistically rejected if it did not also include a multivariate time series component that can capture the fact that the c.r.f. return dynamics are not iid but in fact are dependent across time. We now proceed to explain how SSA can be embedded in a dynamic risk factor model and therefore how the weaknesses mentioned above can be overcome. We note that the dynamic risk factor model is not intended to replace the non-probabilistic model of (3). Indeed it is quite possible the portfolio manager likes to think in terms of the risk factors $\mathbf{f}_{1:l,t+1}$ and would be reluctant to see these replaced by alternative risk factors. The goal here then is to *embed* (3) in a dynamic risk factor model as in (2).

4. A Dynamic Factor Model-Based Approach to Scenario Analysis

In order to embed the SSA approach within a dynamic factor model we need to be able to perform the following steps:

1. Select and estimate a multivariate times series or state-space model for the c.r.f. returns $\mathbf{f}_{t+1}$. We need to be able to handle both observable and latent factors.
2. Specify a factor model (2) for the risk factor changes $\boldsymbol{\Delta}\mathbf{x}_t$.
3. Simulate samples of $\boldsymbol{\epsilon}_{t+1}$ and $\mathbf{f}_{t+1} \mid (\mathcal{F}_t,\, \mathbf{f}_{\mathbf{s},t+1})$.
4. Compute the portfolio P&L (1) for each simulated sample from Step 3. Given these sample P&L's we can estimate the expected P&L for that scenario as well as any other quantities of interest, e.g. a VaR or CVar for that scenario.

Together Steps 1 and 2 enable us to estimate the joint distribution of the common factor returns conditional on time t information. Specifically, they enable us to estimate the distribution π_{t+1} where

$$\mathbf{f}_{t+1} \mid \mathcal{F}_t \sim \pi_{t+1}. \tag{5}$$

We assume $\mathcal{F}_t$ includes the time series of risk factor changes $\boldsymbol{\Delta}\mathbf{x}_0, \ldots, \boldsymbol{\Delta}\mathbf{x}_{t-1}$, as well as the time series of *observable* c.r.f. returns. Step 3 then enables us to generate samples from the distribution of the risk factors $\boldsymbol{\Delta}\mathbf{x}_t$ conditional on $\mathcal{F}_t$ and the scenario $\mathbf{f}_{\mathbf{s},t+1}$. Given these samples, Step 4 is a matter of computing the portfolio P&L for each sample and we assume this step is a straightforward task so that any pricing models required to compute $\Delta V_t(\boldsymbol{\Delta}\mathbf{x}_t)$ given $\boldsymbol{\Delta}\mathbf{x}_t$ are available and easy to implement.

As mentioned above, c.r.f.'s can be either observable or latent. Observable c.r.f.'s might include market indices such as the S&P 500 or the Eurostoxx 50 index, foreign exchange rates, index CDS rates, commodity prices etc. C.r.f. returns that are latent, however, can only be inferred or estimated from other observable data such as the $\mathbf{\Delta x}_t$'s or observable c.r.f.s. Examples of latent c.r.f.'s might include factors that drive the implied volatility surface of the S&P 500, for example. A popular specification would include three c.r.f.'s that drive parallel shifts, term-structure shifts and skew shifts in the implied volatility surface, respectively. Note that such shifts are never observable and can only be inferred from the changes (the $\mathbf{\Delta x}_t$'s) in the implied volatilities of S&P 500 options of various strike-maturity combinations. Another example of latent c.r.f.'s would be the factors that are motivated by a principal components analysis (PCA) of the returns on US Treasuries of various maturities. While there may be twenty or more maturities available, a PCA analysis suggests that changes in the yield curve are driven by just three latent factors representing, in order of importance, a parallel shift in the yield curve, a steepening / flattening of the yield curve, and a change in curvature of the yield curve, respectively.

Because most settings have one or more latent c.r.f.'s our main focus will be on the use of state-space models to tackle steps 1 to 3. We begin with the case where all c.r.f.'s are latent.

4.1. State-Space Modeling of the Common Factor Returns

One way to proceed when all c.r.f. returns are latent is to simply construct point estimates of the latent factors by solving for $k = 1, \ldots, t$ an MLE problem[4] of the form

$$\max_{\mathbf{f}_k \in \mathbb{R}^m} \mathbb{P}_\epsilon(\mathbf{\Delta x}_{k-1} - \mathbf{B}\mathbf{f}_k) \tag{6}$$

where $\mathbb{P}_\epsilon(\cdot)$ is the PDF of $\boldsymbol{\epsilon}_k$ from (2). Let $\hat{\mathbf{f}}_k$ denote the optimal solution to (6). We could then take the $\hat{\mathbf{f}}_k$'s to be observable risk factors and use them to estimate the distribution π_{t+1} of $\mathbf{f}_{t+1} \mid \mathcal{F}_t$. This is clearly sub-optimal, however, as the estimation of the $\mathbf{f}_k$'s ignores the temporal dependence in their dynamics. Moreover, by treating $\hat{\mathbf{f}}_t$ as the true value of $\mathbf{f}_t$ (rather than just a noisy point estimate), we are underestimating the conditional uncertainty in $\mathbf{f}_{t+1}$ when we use these point estimates to estimate π_{t+1}.

Our second and preferred approach overcomes these issues by defining a state-space model for the unobservable common factors then and treating $\mathbf{\Delta x}_0, \ldots, \mathbf{\Delta x}_{t-1}$ as noisy observations of the underlying states $\mathbf{f}_1, \ldots, \mathbf{f}_t$. For example, we could model the unobservable common factor returns via an auto-regressive stochastic process of the form

$$\mathbf{f}_{t+1} = \mathbf{G}\mathbf{f}_t + \boldsymbol{\eta}_{t+1} \tag{7}$$

for some matrix $\mathbf{G} \in \mathbb{R}^{m\times m}$ and where the process innovation terms $\boldsymbol{\eta}_t \in \mathbb{R}^m$ are assumed to have zero mean and constant covariance matrix $\mathbf{\Sigma}_\eta$. The initial state $\mathbf{f}_0$ is assumed to follow some probability distribution π_0. The hidden-state process (7) together with the observable risk factor changes $\mathbf{\Delta x}_t$ from the factor model in (2) now form a state-space model.

As before, our goal is to estimate π_{t+1}, the distribution of $\mathbf{f}_{t+1} \mid \mathcal{F}_t$, where $\mathcal{F}_t$ now only includes the history of observations $\mathbf{\Delta x}_{0:t-1} \coloneqq \{\mathbf{\Delta x}_0, \ldots, \mathbf{\Delta x}_{t-1}\}$. Note that if we are able to obtain the filtered probability distribution $\mathbb{P}(\mathbf{f}_t \mid \mathbf{\Delta x}_{0:t-1})$, then (7) implies we can obtain π_{t+1} as the convolution of the two random

[4] As an alternative to (6) we could obtain the point estimate $\hat{\mathbf{f}}_t$ by solving a cross-sectional regression problem.

variables $\mathbf{G}\mathbf{f}_t \mid \mathcal{F}_t$ and $\boldsymbol{\eta}_{t+1}$. Suppose for example, that π_0 and both process innovations $\boldsymbol{\eta}_{t+1}$ in (7) and $\boldsymbol{\epsilon}_{t+1}$ in (2) are all Gaussian. Then the filtered distribution $\mathbf{f}_{t+1} \mid \mathcal{F}_t$ is also Gaussian and its mean vector and covariance matrix can be calculated explicitly via the Kalman Filter [21]. In this case π_{t+1} would then also be Gaussian.

For non-Gaussian state-space models, however, obtaining the posterior probability exactly is generally an intractable problem although there are many tractable approaches that can be used to approximate the distribution of $\mathbf{f}_{t+1} \mid \mathcal{F}_t$. The Extended Kalman Filter and the Unscented Kalman Filter [35] can be used for non-linear Gaussian state space models, for example. More generally particle filters [20] or MCMC [34] could also be used to approximate the filtered distribution for non-Gaussian state-space models. Particle filters suffer from the curse of dimensionality, however, while MCMC is computationally expensive. Nonetheless implementing an MCMC or particle filter (in the lower dimensional setting) for non-linear / non-Gaussian state-space models should not be too demanding given modern computing power.

As an alternative to computing or approximating the filtered distribution $\mathbb{P}(\mathbf{f}_t \mid \boldsymbol{\Delta}\mathbf{x}_{0:t-1})$, we could simply compute its posterior mean $\mathbb{E}[\mathbf{f}_t \mid \boldsymbol{\Delta}\mathbf{x}_{0:t-1}]$ or its maximum a posteriori (MAP) estimate. Then, using $\hat{\mathbf{f}}_t$ as an approximation to the actual realization of $\mathbf{f}_t$, we can approximate π_{t+1} as the distribution of $\mathbf{G}\hat{\mathbf{f}}_t + \boldsymbol{\eta}_{t+1}$, i.e., the right-hand-side of (7), which would simply be the distribution of $\boldsymbol{\eta}_{t+1}$ shifted to have mean $\mathbf{G}\hat{\mathbf{f}}_t$. While this neglects the uncertainty in our estimation of $\mathbf{f}_t$ this is often a second-order issue relative to obtaining the correct mean of π_{t+1}.

4.2. Modeling Both Observable and Unobservable Common Factor Returns

Situations in which there are a combination of observable and latent common factor returns are not uncommon. For example, in an S&P 500 options portfolio a scenario would typically include stresses to some combination of the S&P 500 (observable) and parallel, skew or term structure shifts (latent) in the S&P 500's implied volatility surface. In this case, the challenge is to construct a multivariate state-space / time series model that can simultaneously accommodate observable and latent c.r.f. returns. While there may be many ways to tackle this modeling problem, one obvious approach is to assume all of the c.r.f.'s are latent but that the noisy signals for a subset of them (the observable ones) are essentially noiseless.

To make this more precise, we assume we have m^o observable and m^u latent common factors so that the factor model (2) can then be written as

$$\boldsymbol{\Delta}\mathbf{x}_t = \mathbf{B}^o \mathbf{f}^o_{t+1} + \mathbf{B}^u \mathbf{f}^u_{t+1} + \boldsymbol{\epsilon}_{t+1} \tag{8}$$

where $\mathbf{B}^o \in \mathbb{R}^{n \times m^o}$ and $\mathbf{B}^u \in \mathbb{R}^{n \times m^u}$ are the factor loadings matrices for the observable and latent common factors $\mathbf{f}^o_{t+1}$ and $\mathbf{f}^u_{t+1}$, respectively. Our objective is to estimate π_{t+1}, the probability distribution of $\mathbf{f}_{t+1} \mid \mathcal{F}_t$, where $\mathcal{F}_t$ now corresponds to the σ-algebra generated by the history of risk factor changes $\boldsymbol{\Delta}\mathbf{x}_{0:t-1}$ and of the observable common factor returns $\mathbf{f}^o_{1:t}$. We define the $n^o := n + m^o$ dimensional vector

$$\mathbf{y}_t := \begin{bmatrix} \boldsymbol{\Delta}\mathbf{x}_t \\ \mathbf{f}^o_{t+1} \end{bmatrix}$$

which we treat as the time $t+1$ observations vector. The model's latent state variables at time $t+1$ are given by the $m := m^o + m^u$ dimensional vector

$$\mathbf{f}_{t+1} := \begin{bmatrix} \mathbf{f}_{t+1}^o \\ \mathbf{f}_{t+1}^u \end{bmatrix}.$$

We now assume the observation dynamics satisfy

$$\mathbf{y}_t = \begin{bmatrix} \boldsymbol{\Delta}\mathbf{x}_t \\ \mathbf{f}_{t+1}^o \end{bmatrix} = \begin{bmatrix} \mathbf{B}^o & \mathbf{B}^u \\ \mathbf{I}_{m^o} & \mathbf{0}_{m^{ou}} \end{bmatrix} \begin{bmatrix} \mathbf{f}_{t+1}^o \\ \mathbf{f}_{t+1}^u \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_{t+1} \\ \mathbf{0}_{m^o} \end{bmatrix} \qquad t = 0, 1, \ldots \tag{9}$$

where $\mathbf{I}_{m^o}$ is the $m^o \times m^o$ identity matrix, $\mathbf{0}_{m^{ou}}$ is the $m^o \times m^o$ matrix of zeros and $\mathbf{0}_{m^o}$ is an $m^o \times 1$ vector of zeros. We can again assume latent state dynamics of the form given in (7). Since (7) and (9) form a state-space model, we can fit the model and estimate π_{t+1} using the various approaches described above. For instance, assuming $\boldsymbol{\epsilon}_{t+1}$ and $\boldsymbol{\eta}_{t+1}$ to be normally distributed, we could use the EM algorithm to estimate the parameters of the state-space model (7) and (9) using historical data. The Kalman Filter can then be employed to obtain the filtered probability distribution $\mathbb{P}(\mathbf{f}_t \mid \mathbf{y}_{0:t-1})$ for any sequence of observations $\mathbf{y}_{0:t-1}$. Finally, we can then obtain π_{t+1} exactly as the sum of two normal random vectors $\mathbf{G}\mathbf{f}_t \mid \mathcal{F}_t$ and $\boldsymbol{\eta}_{t+1}$, which of course is also normal.

We will use a model of the form (7) and (9) in Section 6 where we consider portfolios containing options and futures on the S&P 500 index. One obvious weakness with this model is that it fails to capture some of the well-known behavior of S&P 500 returns (and equity returns more generally). For example, we cannot capture volatility clustering using such a model. In Section 6 we will overcome this by embedding a GARCH model for the S&P 500 returns within a state-space model of the form (7) and (9). We will also provide empirical evidence in Section 7 that this GARCH / state-space model combination is a clear improvement over the simple state-space model of (7) and (9). More generally, it should be possible to move beyond linear state-space models via the use of MCMC, particle filters etc. to build more realistic models of c.r.f. return dynamics.

5. Evaluating the Performance of SSA

The objective of the dynamic factor model-based scenario analysis (hereafter DFMSA) is to compute

$$\Delta V_t^{\text{dfm}}(\mathbf{c}) := \mathbb{E}_t[\Delta V_t(\mathbf{f}_{t+1}, \boldsymbol{\epsilon}_{t+1}) \mid \mathbf{f}_{\mathbf{s},t+1} = \mathbf{c}] \tag{10}$$

where $\mathbf{c}$ denotes the levels of the stressed factors in the given scenario, and $\mathbb{E}_t[\cdot] := \mathbb{E}[\cdot \mid \mathcal{F}_t]$ denotes an expectation taken with respect to the distribution π_{t+1}. It's clear we have to be able to compute or simulate from the distribution of $\mathbf{f}_{t+1} \mid (\mathcal{F}_t, \mathbf{f}_{\mathbf{s},t+1} = \mathbf{c})$ in order[5] to calculate the conditional expectation in (10). Since π_{t+1} is the true conditional distribution of $\mathbf{f}_{t+1} \mid \mathcal{F}_t$, we know that $\Delta V_t^{\text{dfm}}(\mathbf{c})$ is the best estimate (in the mean-squared error sense) of the scenario P&L. We can therefore calculate the error obtained from following the SSA approach for a given scenario $\mathbf{c}$ as

$$\mathrm{E}_t^{\text{abs}}(\mathbf{c}) := |\Delta V_t^{\text{dfm}}(\mathbf{c}) - \Delta V_t^{\text{ss}}(\mathbf{c})| \tag{11}$$

[5]We also note that the conditional distribution of $\boldsymbol{\epsilon}_{t+1} \mid (\mathcal{F}_t, \mathbf{f}_{\mathbf{s},t+1} = \mathbf{c})$ (where $\boldsymbol{\epsilon}_{t+1}$ is given in (2)) is equal to its unconditional distribution since $\boldsymbol{\epsilon}_{t+1}$ is independent of $\mathcal{F}_t$ and $\mathbf{f}_{t+1}$ by assumption.

where $\Delta V_t^{\text{ss}}(\mathbf{c})$ denotes the estimated scenario P&L at time t according to the SSA approach. We must of course acknowledge that the error in (11) is somewhat misleading in that it assumes our dynamic factor model is indeed the correct model that governs the real-world security price dynamics. Moreover, it seems reasonable to assume there is some dynamic factor model that governs the real-world security price dynamics and that we have been able to construct a reasonable approximation to it. After all, if our model is not a reasonably good approximation to it, then it would be rejected by one or more of the statistical tests that are discussed in Section 7. As such, we feel it is reasonable to take (11) as a ballpark estimate of the error than can arise from adopting the SSA approach.

We can also provide a partial decomposition of the error in (11) by calculating an alternative scenario P&L that is given by

$$\Delta V_t^{\text{alt}}(\mathbf{c}) := \Delta V_t(\mathbf{B}\boldsymbol{\mu}_t^c) \tag{12}$$

where $\mathbf{B}$ is the factor loadings matrix of the factor model (2) and

$$\boldsymbol{\mu}_t^c := \mathbb{E}_t[\mathbf{f}_{t+1} \mid \mathbf{f}_{\mathbf{s},t+1} = \mathbf{c}]. \tag{13}$$

This alternative scenario P&L estimator (suggested in point #3 from Section 3.2) goes beyond the SSA approach by using the expected value of the c.r.f. returns conditional on the scenario to estimate the risk factor changes $\boldsymbol{\Delta}\mathbf{x}_t$ via the factor model (2), i.e. by setting $\boldsymbol{\Delta}\mathbf{x}_t = \mathbf{B}\boldsymbol{\mu}_t^c$. This leads to the alternative estimated scenario P&L in (12). Note that the alternative scenario P&L $\Delta V_t^{\text{alt}}(\mathbf{c})$ will in general differ from[6] and be less accurate than $\Delta V_t^{\text{dfm}}(\mathbf{c})$ as defined in (10). We can then decompose the error in (11) by

$$\begin{aligned} \mathrm{E}_t^{\text{abs}}(\mathbf{c}) &= |\Delta V_t^{\text{dfm}}(\mathbf{c}) - \Delta V_t^{\text{ss}}(\mathbf{c})| \\ &= |\Delta V_t^{\text{dfm}}(\mathbf{c}) - \Delta V_t^{\text{alt}}(\mathbf{c}) + \Delta V_t^{\text{alt}}(\mathbf{c}) - \Delta V_t^{\text{ss}}(\mathbf{c})| \\ &\leq |\Delta V_t^{\text{dfm}}(\mathbf{c}) - \Delta V_t^{\text{alt}}(\mathbf{c})| + |\Delta V_t^{\text{alt}}(\mathbf{c}) - \Delta V_t^{\text{ss}}(\mathbf{c})|. \end{aligned} \tag{14}$$

We note that $|\Delta V_t^{\text{dfm}}(\mathbf{c}) - \Delta V_t^{\text{alt}}(\mathbf{c})|$ gives a measure of the error that results from ignoring the variance in the conditional distribution of the c.r.f. returns and the idiosyncratic error terms. In contrast, $|\Delta V_t^{\text{alt}}(\mathbf{c}) - \Delta V_t^{\text{ss}}(\mathbf{c})|$ provides a measure of the error that results from setting the unstressed common factor returns to zero rather than their conditional expected values. While the sum of these two errors does not equal the true error we see from (14) that their sum does provide an upper bound on this error. In our numerical applications we found that the second term on the r.h.s. of (14), i.e. $|\Delta V_t^{\text{alt}}(\mathbf{c}) - \Delta V_t^{\text{ss}}(\mathbf{c})|$, is considerably more significant than the first term on the r.h.s. and often provides a very good approximation to the true error on the l.h.s. of (14). Of course this may not be the case in general, particularly with portfolios whose P&L is very non-linear in the risk factors $\boldsymbol{\Delta}\mathbf{x}_t$ and where the conditional variance of the non-stressed factors is substantial.

[6] Suppose for example that $\Delta V_t(\cdot)$ is a convex function. Then Jensen's inequality implies $\Delta V_t^{\text{alt}}(\mathbf{c}) = \Delta V_t(\mathbf{B}\boldsymbol{\mu}_t^c) = \Delta V_t(\mathbb{E}_t[\mathbf{B}\mathbf{f}_{t+1} + \boldsymbol{\epsilon}_{t+1} \mid \mathbf{f}_{\mathbf{s},t+1} = \mathbf{c}]) \leq \mathbb{E}_t[\Delta V_t(\mathbf{B}\mathbf{f}_{t+1} + \boldsymbol{\epsilon}_{t+1}) \mid \mathbf{f}_{\mathbf{s},t+1} = \mathbf{c}] = \Delta V_t^{\text{dfm}}(\mathbf{c})$. In this case $\Delta V_t^{\text{alt}}(\mathbf{c})$ would underestimate the estimated scenario P&L when $\Delta V_t(\cdot)$ is convex. Similarly $\Delta V_t^{\text{alt}}(\mathbf{c})$ would overestimate the estimated scenario P&L when $\Delta V_t(\cdot)$ is concave.

5.1. Back-Testing Procedure for Evaluating SSA

In our numerical experiments we will simulate a ground truth model for T periods and for each period compute the SSA error as defined in (11). We can then average these errors across time to get some idea of how poorly (or well) SSA performs in relation to DFMSA. Since the ground truth model will coincide with the dynamic factor-model that we use to perform the scenario analysis, this approach assumes the estimated P&Ls from the DFMSA are correct in the mean-squared error sense. While of course this is optimistic, it does serve to highlight just how inaccurate the P&Ls reported by SSA can be. It is also worth emphasizing that while we assume we know the *structure* of the ground truth model in these back-tests, we still do not get to observe the latent c.r.f.'s. These latent factor returns must be inferred in our back-tests from the risk factor returns, i.e. the $\boldsymbol{\Delta}\mathbf{x}_t$'s, as well as the observable c.r.f. returns. In general, we will also be required to re-estimate the parameters of the model each day within the back-tests rather than simply assuming these parameters are given and known to us.

More specifically, in each of our back-tests we assume we have T days of simulated data. We choose s where $0 < s < T$ to be the size of the rolling window that we will use to re-estimate the model at each time $t \geq s$. Having estimated the dynamic-model's parameters, we then estimate π_{t+1} and use it to estimate the DFMSA P&L $\Delta V_t^{\mathrm{dfm}}$. The SSA P&L ΔV_t^{ss} is also computed at this time. At the end of the back-test we can calculate the average of the back-test P&L for each approach according to

$$\overline{\Delta V}_{\mathrm{dfm}} := \frac{1}{T-s} \sum_{t=s}^{T-1} \Delta V_t^{\mathrm{dfm}} \qquad\qquad \overline{\Delta V}_{\mathrm{ss}} := \frac{1}{T-s} \sum_{t=s}^{T-1} \Delta V_t^{\mathrm{ss}} \tag{15}$$

Comparing $\overline{\Delta V}_{\mathrm{ss}}$ with $\overline{\Delta V}_{\mathrm{dfm}}$ gives a measure of the bias of the SSA approach over the course of the back-test. We can also calculate the mean absolute difference between the estimated SSA P&L and the estimated DFMSA P&L. That is we define

$$\mathrm{E}^{\mathrm{abs}} := \frac{1}{T-s} \sum_{t=s}^{T-1} \left| \Delta V_t^{\mathrm{dfm}} - \Delta V_t^{\mathrm{ss}} \right| \tag{16}$$

as the average error in the P&L estimated by the SSA approach. Of course this error depends on the ground truth model and its parameters as well as the portfolio and scenario under consideration. Our general back-testing procedure is outlined in Algorithm 1 below.

5.2. What Portfolios to Back-Test?

Before proceeding to our numerical experiments, it is worth discussing what kinds of portfolios we have in mind when comparing the SSA approach with the DFMSA approach. For all of the reasons outlined earlier we would argue that, regardless of the portfolio, any scenario analysis ought to be embedded in a dynamic factor model setting. Nonetheless, it stands to reason that certain types of portfolios might show little difference between the scenario P&Ls reported by the SSA and DFMSA approaches, respectively. On the other hand, it is not difficult to imagine settings where the two scenario P&Ls might be very different. For example, consider a setting with securities whose daily P&L's are non-linear functions of their risk factor changes and where some of the c.r.f. returns are at least moderately[7] dependent. Consider now a portfolio

[7]The assumption that some of the c.r.f. returns might display moderate dependence is not a strong assumption since even uncorrelated c.r.f. returns can display moderate dependence. Suppose for example that the c.r.f. returns have a joint multivariate t distribution with ν degrees-of-freedom. These factor returns can be uncorrelated and yet still have extreme tail

Algorithm 1 back-testing to Estimate Average SSA Error for a Given Scenario and Ground-Truth Model

Input: s, T, K, **gmodel**, $\mathbf{C}, \mathbf{c}$ ▷ s = # periods in rolling window for model training
▷ T = # periods in back-test horizon
▷ K = # of samples used to estimate factor model-based scenario P&L
▷ **gmodel** is the ground-truth model
▷ $\mathbf{c}$, $\mathbf{s}$ define the scenario.

Generate $\mathbf{f}_0$ from **gmodel**
for $t \leftarrow 0$ to $T-1$ **do**
  Generate $(\mathbf{f}_{t+1}, \mathbf{\Delta x}_t) \mid \mathbf{f}_t$ from **gmodel**
  if $t \geq s$ **then**
    Estimate DFM parameters
    Estimate π_{t+1} from $(\mathbf{f}^o_{(t-s):t}, \mathbf{\Delta x}_{(t-s):(t-1)})$ ▷ $\mathbf{f}^o_{t-s:t}$ are observable
    for $k \leftarrow 1$ to K **do** ▷ K = # of Monte-Carlo samples
      Generate $\mathbf{f}^{(k)}_{t+1} \mid (\mathcal{F}_t, \mathbf{f}_{\mathbf{s},t+1} = \mathbf{c})$ and $\boldsymbol{\epsilon}^{(k)}_{t+1}$ to obtain $\mathbf{\Delta x}^{(k)}_t$
      Compute scenario P&L $\Delta V_t(\mathbf{\Delta x}^{(k)}_t)$
    end for
    Compute $\Delta V^{\text{dfm}}_t := \sum_{k=1}^{K} \Delta V_t(\mathbf{\Delta x}^{(k)}_t)/K$ ▷ Estimated scenario P&L
    Compute ΔV^{ss}_t ▷ SSA P&L obtained by setting $\boldsymbol{\epsilon}_{t+1}$, non-stressed c.r.f. returns to $\mathbf{0}$
    Compute $\mathrm{E}^{\text{abs}}_t := \left|\Delta V^{\text{dfm}}_t - \Delta V^{\text{ss}}_t\right|$
  end if
end for
Compute $\overline{\Delta V}_{\text{dfm}}, \overline{\Delta V}_{\text{ss}}$ and E^{abs} as defined in (15) and (16)

Output: $\overline{\Delta V}_{\text{dfm}}, \overline{\Delta V}_{\text{ss}}$ and E^{abs}

that was designed to be: (i) neutral to the subset of c.r.f.'s that are stressed in scenarios and (ii) highly exposed to the c.r.f. returns that are never stressed in any of the scenarios. If some of the non-stressed c.r.f. returns are conditionally dependent on some of the stressed c.r.f. returns then such a portfolio should result in very different scenario P&Ls for the SSA and DFMSA approaches.

For an adversarial example, let $\mathbf{f_e}$ where $\mathbf{e} \subset \{1, \ldots, m\}$ denote the subset of the c.r.f.'s to which the p.m. wants to be exposed. It's possible for example that the p.m. has a strong view regarding the direction of $\mathbf{f_e}$ over the short term and wishes to trade on that view. Similarly, let $\mathbf{f_n}$ where $\mathbf{n} \subset \{1, \ldots, m\}$ denote the set of c.r.f.'s to which the trader is required to be neutral according to the risk-management team. We assume the p.m. computes scenario P&Ls using the DFMSA approach while the risk team uses the SSA approach. The p.m. can then easily construct a risky portfolio that gives her the desired exposure to $\mathbf{f_e}$ but that appears to have little risk according to the risk management team's perspective. If some of the c.r.f. returns in $\mathbf{f_n}$ are dependent (conditional on the scenario) with some of the c.r.f. returns in $\mathbf{f_e}$ then this portfolio should result in very different scenario P&Ls for the SSA and DFMSA approaches. In Appendix A we outline a simple linear programming approach for constructing such portfolios and we will use these portfolios in our numerical experiments of Section 6 and Appendix D.

We also note that this setting is not at all contrived since it is quite possible for a p.m. to have a strong view on a less important risk factor which may not be a risk-factor considered by the risk-management team. Less generously, it may be the case that the p.m. is incentivized to take on a lot of risk regardless of whether or not he / she has a view justifying this risk-taking. Regardless of the p.m.'s motivation, the use of SSA instead of DFMSA can lead to very misleading scenario P&L's.

dependence [23]. As a result the distribution of these factors conditional on an extreme scenario can display strong dependence.

6. An Application to an S&P Options Portfolio

In this application we consider a p.m. that can invest in European call and put options on the S&P 500 index as well as in the index itself. As is standard market practice, we will use the Black-Scholes formula to price these options. To value the portfolio, we will therefore[8] assume the vector of risk factor changes $\mathbf{\Delta x}_t$ consists of the daily log-return of the S&P 500 together with daily changes in the implied volatilities of specific strike-maturity combinations.

In order to capture well-known features such as volatility clustering in the S&P 500 dynamics, we will not use the daily log-return of the S&P 500 directly in $\mathbf{\Delta x}_t$. Instead we propose to use the innovations process resulting from a GARCH model[9] fitted to the S&P 500 log-returns. These innovations have the property of having a constant volatility (which we can set to 1) and can be therefore thought of as the *devolatilized* S&P 500 log-returns. More precisely, we assume that $r_t := \log(S_{t+1}/S_t)$ follows a GARCH$(1,1)$ model given by

$$r_t = \sigma_t z_t, \qquad t = 1, 2, \ldots \tag{17}$$

where the innovations z_t form a process of i.i.d. random variables with zero mean and unit variance, and where σ_t is the *conditional* volatility of r_t, and is given by

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \qquad t = 1, 2, \ldots \tag{18}$$

with $\alpha_0 > 0$ and $\alpha_1, \beta_1, \sigma_0 \geq 0$. From (17) and (18) we note that the conditional volatility of the r_t's, given by σ_t, is dynamic through time and can capture volatility clustering through its non-negative dependence on the volatility of the prior period. We then take $z_t = \log(S_{t+1}/S_t)/\sigma_t$ to be the first element of the vector of risk factor changes $\mathbf{\Delta x}_t$.

We let $I_t(\xi, \tau)$ denote the implied volatility at time t of a European option with time-to-maturity τ and option moneyness $\xi := K/S_t$ where K denotes the option strike and S_t the time t price of the S&P 500. We assume that on each day we can observe the implied volatility surface at a finite set of moneyness-maturity pairs $\{(\xi_1, \tau_1), \ldots, (\xi_{n-1}, \tau_{n-1})\}$. For a fixed pair (ξ, τ), we denote the change in implied volatility from t to $t+1$ by

$$\Delta I_t(\xi, \tau) := I_{t+1}(\xi, \tau) - I_t(\xi, \tau). \tag{19}$$

The vector of risk factors changes is then given by the n-dimensional vector

$$\mathbf{\Delta x}_t := (z_t, \Delta I_t(\xi_1, \tau_1), \ldots, \Delta I_t(\xi_{n-1}, \tau_{n-1}))^\top \tag{20}$$

where the moneyness-maturity pairs in (20) cover the distinct moneyness-maturity combinations of the options in the market.

Our dynamic factor model will consist of four c.r.f.'s. We will take the innovations z_t (obtained from

[8]We assume the risk-free rate of interest and dividend yield remain constant throughout and therefore do not model risk factors associated with them. This is typical for an equity options setting unless the p.m. wishes to trade with a specific view on dividends. We also acknowledge that in practice one trades futures on the S&P 500 index rather than the index itself. Given the assumption of a constant risk-free rate and dividend yield, there is essentially no difference in assuming we can trade the index itself, however, and so we will make that assumption here.

[9]GARCH models are a well-established class of time-series models that can model empirically observed behavior such as volatility clustering in equity markets. [23] and [34] can be consulted for a more detailed exposition as well as related references.

the GARCH model of the S&P 500's daily log-returns) to be the first c.r.f. and of course this is observable. The other $m = 3$ c.r.f.'s will be latent factors that drive changes in the implied volatility surface, specifically parallel[10] changes in the surface, a steepening / flattening of the volatility skew, and a steepening / flattening of the term structure. As our model will contain both observable and latent c.r.f.s we will proceed as discussed in Section 4.2 and use a linear Gaussian state-space model. In particular, we will use a slightly modified version of (9) and define

$$\mathbf{\Delta x}_t = \begin{bmatrix} 1 \\ \mathbf{b}^o \end{bmatrix} f^o_{t+1} + \begin{bmatrix} \mathbf{0}_3^\top \\ \mathbf{B}^u \end{bmatrix} \mathbf{f}^u_{t+1} + \begin{bmatrix} 0 \\ \boldsymbol{\epsilon}_{t+1} \end{bmatrix} \tag{21}$$

where $f^o_{t+1} := z_t$ is the observable c.r.f. which coincides with the first component of $\mathbf{\Delta x}_t$, $\mathbf{f}^u_{t+1} \in \mathbb{R}^3$ denotes the vector of latent c.r.f.s and $\mathbf{0}_3 \in \mathbb{R}^3$ denotes the zero vector. The factor loadings for the observable and latent c.r.f.s are denoted by $\mathbf{b}^o \in \mathbb{R}^{n-1}$ and $\mathbf{B}^u \in \mathbb{R}^{(n-1)\times 3}$, respectively. The i^{th} element of $\mathbf{b}^o$ indicates how a shock to z_t affects the implied volatility $\Delta I_t(\xi_i, \tau_i)$. The matrix $\mathbf{B}^u$ is constructed to model the aforementioned $m = 3$ types of stresses on the implied volatility surface. Specifically (and recalling that the $(i+1)^{st}$ component of $\mathbf{\Delta x}_t$ is $\Delta I_t(\xi_i, \tau_i)$), we assume

$$\Delta I_t(\xi_i, \tau_i) = b^o_i f^o_{t+1} + \left(\frac{1}{\sqrt{\tau_i}}\right) f^u_{1,t+1} + (1 - \xi_i)\, f^u_{2,t+1} + \ln(2\tau_i) f^u_{3,t+1} + \epsilon_{i,t+1}, \qquad \text{for } i = 1, \ldots, n-1 \tag{22}$$

where b^o_i, $f^u_{i,t+1}$ and $\epsilon_{i,t+1}$ denote the i^{th} elements of $\mathbf{b}^o$, $\mathbf{f}^u_{t+1}$ and $\boldsymbol{\epsilon}_{t+1}$, respectively, in (21). Comparing (21) and (22), we see that $b_{i,1} := 1/\sqrt{\tau_i}$, $b_{i,2} := 1 - \xi_i$ and $b_{i,3} := \ln(2\tau_i)$ where $b_{i,j}$ is the $(i,j)^{th}$ element of $\mathbf{B}^u$.

A few comments on (22) are now in order. It is well known (see for example Natenberg [24]) that when volatility rises (falls), the implied volatility of long-term options rises (falls) less than the implied volatility of short-term options. This empirical observation has led to the commonly used "square-root-of-time" rule whereby the relative difference in implied volatility changes for options with the same moneyness but different maturities is in proportion to the square-roots of their relative maturities. We model this rule via the factor loadings for the parallel-shift c.r.f. Suppose, for example, there is a $f^u_{1,t+1} = 1$ volatility point shock to the parallel-shift c.r.f. Then the $f^u_{1,t+1}/\sqrt{\tau_i}$ term in (22) implies that the implied volatility of 1-year options would increase by 1 point exactly, whereas the implied volatility of a 1-month option would increase by $1/\sqrt{1/12} \approx 3.46$ points.

The second latent c.r.f. is used to drive changes in the implied volatility skews[11] in the surface. We use the so-called "sticky-moneyness" rule which assumes that, for a given maturity, the implied volatility is a univariate function of the moneyness $\xi = K/S$. The "sticky-moneyness" rule that we adopt in (22) can be

[10] We acknowledge that the absence of arbitrage imposes restrictions on the magnitude of permissable c.r.f. stresses. For example, Rogers [31] has shown that the implied volatility surface cannot move in parallel without introducing arbitrage opportunities. Indeed it is well known that moves in the implied volatilities are more likely to follow a "square-root-of-time" rule and we will model this below with our first latent c.r.f. For another example, it is also well-known that that the volatility skew at any fixed maturity cannot become too steep without introducing arbitrage. We don't explicitly rule out scenarios that allow for arbitrage but note that such scenarios would have to be very extreme indeed. Moreover, it is easy to check a given scenario for arbitrage and so ruling out such scenarios would be very straightforward.

[11] An implied volatility skew is the cross-section of the implied volatility surface that we obtain when we hold the time-to-maturity fixed. There is therefore a different skew for each time-to-maturity. There are various skew models in the literature and we refer the interested reader to the work of Derman and Miller [12] who describe some of these models.

loosely motivated by first assuming

$$I_t(\xi,\tau) = I_0(1,\tau) - \beta_t\,(\xi - 1) \tag{23}$$

where $I_0(1,\tau)$ is the implied volatility of an at-the-money option, i.e. with $\xi = 1$, with maturity τ at some initial time $t = 0$, and where β_t determines the slope of the skew at time t. The model (23) implies the implied volatility for at-the-money options (for which $\xi = 1$) remains constant for a given maturity τ, and that changes in implied volatility are given by

$$\Delta I_t(\xi,\tau) = -\Delta\beta_t\,(\xi - 1) \tag{24}$$

where $\Delta\beta_t := \beta_{t+1} - \beta_t$ defines the change in skew (or slope) of the implied volatility. We account for this skew behavior in our factor model (21) by taking $\Delta\beta_t$ to be the c.r.f. $f^u_{2,t+1}$ and setting the corresponding factor loadings to $(1-\xi)$. Then if $f^u_{2,t+1} > 0$, for example, the implied volatilities of options with moneyness < 1 (> 1) would increase (decrease) thereby resulting in the steepening of the skew for any given maturity τ. Similarly a shock $f^u_{2,t+1} < 0$ would result in a flattening of the skew.

The third c.r.f. $f^u_{3,t+1}$ models changes to the term-structure of implied volatility for any given level of moneyness. The loading term $\ln(2\tau_i)$ means that a positive shock to $f^u_{3,t+1}$ would leave 6-month volatilities unchanged, but would increase (decrease) the volatilities of options with longer (shorter) maturities thereby resulting in the flattening of an inverted term structure or steepening of an already upward sloping term structure. We note that the parallel shift c.r.f. also affects the term structure due to the square-root-of-time rule. However, including the term structure c.r.f. enriches the dynamics of the volatility surface model as it allows for a broader variety of systematic moves, i.e. moves driven only by the c.r.f.s.

Finally, we note that at this point we are neither arguing for or against the specific model of (21) and (22) together with (2). Whether or not the model would work well in practice would depend on its ability to pass the various statistical tests that we describe and implement in Section 7. We will see there that while the model is not perfect, it is a significant improvement over a more naive version of the model where the S&P 500 log-returns (rather than the devolatilized returns) are taken to be the first component of the c.r.f. returns.

6.1. Model Calibration

We obtained implied volatility data on the S&P 500 for the period January 2006 through August 2013 from the OptionMetrics IVY database. In particular, we used the daily implied volatility data that OptionMetrics provide for various delta-maturity[12] combinations. We transformed the data to moneyness-maturity coordinates using a non-parametric Nadaraya-Watson estimator based on an independent bivariate Gaussian kernel [16]. We can therefore obtain the implied volatilities on any given day for the fixed set of moneyness-maturity pairs given by the cross-product of

$$\xi \in \Omega_\xi := \{0.8,\, 0.9,\, 0.95,\, 1.0,\, 1.05,\, 1.1,\, 1.2\} \quad \text{and} \quad \tau \in \Omega_\tau := \{1/12,\, 2/12,\, 3/12,\, 6/12,\, 1\}\,, \tag{25}$$

where the time-to-maturity τ is measured in years.

[12] Roughly speaking, they build an implied volatility surface based on each day's closing prices (of the S&P 500 and its traded options) and then use this surface to read off volatilities for the various delta-maturity combinations.

Using the log-returns of the S&P 500 index for the same period, we fit the GARCH$(1,1)$ model[13] and obtain the resulting *observable* time-series of innovations z_t. We then used the z_t's together with the implied volatility data to fit the linear Gaussian state-space model of (21) and (22) via the EM algorithm. We take our ground-truth model to be the resulting fitted model. The parameters of this ground-truth model are given in Appendix B.

We assume our portfolio can contain the S&P 500 index, at-the-money and out-of-the-money call options with moneyness ($\xi = K/S$) in the set $\{1.00, 1.025, 1.05, 1.075, 1.10, 1.15\}$ and out-of-the-money put options with moneyness in the set $\{0.85, 0.90, 0.925, 0.95, 0.975\}$. The options are assumed to have maturities in the set $\Omega'_\tau := \{i/12 : i = 1, \ldots, 12\}$ so there are a total of $N = 133$ securities in the universe. Each option is priced using the Black-Scholes formula and so we interpolate the implied volatility surface as necessary to obtain the implied volatility for certain moneyness-maturity pairs that are not explicitly modeled. As was the case in Section D.1, we again assume that a risk-free cash account is also available. The cash account is the $(N+1)^{st}$ security and each day it returns 0% w.p. 1.

On each day of our back-test, we construct a portfolio using the LP approach as described in Section 5.2 and Appendix A. We consider a p.m. who on each day t believes (i) the devolatized S&P 500 log-return will be[14] -2 and (ii) the parallel shift c.r.f. will increase by 1 volatility point. We note from (21) that a 1 volatility point increase in the parallel shift c.r.f. would translate to a $1/\sqrt{\tau}$ volatility points increase for options with maturity τ, assuming the idiosyncratic noise and other c.r.f. returns were zero. For example, a 1-month option would then see a $1/\sqrt{1/12} = 3.46$ volatility points increase. These anticipated movements correspond to -2 and +2 standard deviation moves in the two c.r.f.s, respectively.

We assume the p.m. can take short positions on any of the securities except for the cash account so that $0 \le w_{N+1} \le 1$. We also assume that $-0.3 \le w_i \le 0.3$ for $i = 1, \ldots, N$ so that the risk in any one security is limited. We also have the budget constraint $\sum_{i=1}^{N+1} w_i = 1$. In addition to these constraints, we assume that the risk-management desk requires portfolios to be kept "neutral" with respect to a given set of scenarios involving joint stresses to pairwise combinations of the first three c.r.f. returns, i.e., the market, the parallel shift and skew c.r.f. returns. Neutrality to a given scenario is defined as having the portfolio SSA P&L under that scenario to be within $\pm\alpha = 2\%$ of the initial portfolio value. The given scenarios are the elements of the cross-product of $\Omega_{\text{Mkt}} \times \Omega_{\text{ParallelShift}}$ or $\Omega_{\text{Mkt}} \times \Omega_{\text{Skew}}$, where

$$\begin{aligned} \Omega_{\text{Mkt}} &:= \{-3, -2, -1, 0, 1, 2, 3\} \\ \Omega_{\text{ParallelShift}} &:= \{-1.5,\ -1.0,\ -0.5,\ 0.0,\ 0.5,\ 1.0,\ 1.5\} \\ \Omega_{\text{Skew}} &:= \{-9,\ -6,\ -3,\ 0,\ 3,\ 6,\ 9\}. \end{aligned} \tag{26}$$

The values in $\Omega_{\text{Mkt}}, \Omega_{\text{ParallelShift}}$ and Ω_{Skew} were calibrated to be approximately $0, \pm1, \pm2$ and ±3 standard deviations of the z_t, parallel shift and skew c.r.f. returns, respectively. The units of $\Omega_{\text{ParallelShift}}$ and Ω_{Skew} are volatility points. Recall that the magnitude of these movements in the c.r.f.s is given by the corresponding columns of $\mathbf{b}^o$ and $\mathbf{B}^u$ in the factor model (21). For example, a -6 unit move in the skew c.r.f. return translates to a $(1-\xi) \times (-6)$ move in the implied volatility for options with moneyness ξ assuming the idiosyncratic noise and other c.r.f. returns were zero. This translates to a decrease in implied volatility of 0.6 volatility

[13] Standard approaches to fitting GARCH models are Maximum Likelihood Estimation (MLE) and Quasi-MLE methods. For details of these approaches, refer to [23].

[14] Recall that z_t has unit standard deviation and so a fall of -2 in z_t implies via (17) an S&P 500 log-return of $-2\sigma_t$.

points for options with $\xi = 0.9$ and an increase of 0.9 volatility points for options with $\xi = 1.15$. Finally, we also impose the (linear equality) constraints that require the portfolio to be delta, gamma and vega neutral. We impose the latter constraints to allow for the fact that SSA is typically not done in isolation and so it would be typical for any risk manager / p.m. to also know the delta, gamma and vega of the portfolio. By insisting that the portfolio be neutral to the Greeks we are simply making it more difficult for the p.m. to "game" the fact that the scenario constraints are based on SSA rather than the correct DFMSA.

On each day of the back-test we constructed the portfolio and then applied SSA and DFMSA to it using the same scenarios that we used to define the scenario constraints on the portfolio, i.e. scenarios corresponding to shocks in the cross-product of Ω_{Mkt} and $\Omega_{\text{ParallelShift}}$ and Ω_{Mkt} and Ω_{Skew}. We back-tested the model using Algorithm 1 and using the ground-truth model to simulate data for a back-testing horizon of $T = 1,000$ periods. We used an initial training window of size $s = 500$ and then for each $t \in \{s, \ldots, T-1\}$ we fit the GARCH(1,1) model and used the EM algorithm on the observable simulated data $\mathbf{\Delta x}_{(t-s):(t-1)}$ to re-estimate the model parameters $\mathbf{b}^o, \mathbf{G}, \mathbf{\Sigma}_{\boldsymbol{\eta}}$ and $\mathbf{\Sigma}_{\boldsymbol{\epsilon}}$, as well as the parameters of the normal distribution governing the initial state $\mathbf{f}_{t-s}$. Having re-trained the model at time t, we use the Kalman Filter to obtain the distribution of $\mathbf{f}_t \mid \mathbf{\Delta x}_{t-s:t-1}$. We finally obtain π_{t+1} as the distribution of the convolution of $\mathbf{G}\mathbf{f}_t \mid \mathbf{\Delta x}_{(t-s):(t-1)}$ and $\boldsymbol{\eta}_{t+1}$ and simulate samples from π_{t+1} to estimate the scenario P&L's under both SSA and DFMSA approaches.

6.2. Numerical Results

The results of the back-test are displayed in Tables 1 to 3 below. Table 1 contains the average P&L $\overline{\Delta V}_{\text{ss}}$ as estimated by the SSA approach for the same set of scenarios that were used to construct the portfolio. As a result, it is not surprising to see the reported P&L's are all less than 2% in absolute value. The average estimated P&L $\overline{\Delta V}_{\text{dfm}}$ obtained using the DFMSA approach is then reported in Table 2 for the same set of scenarios. It is very clear that the portfolio is (on average) not at all neutral to the various scenarios. For example, when the market c.r.f. return and the parallel shift c.r.f. are shocked by +3 and -1, respectively, the DFMSA approach estimates a loss of 11.9% whereas the SSA approach estimates a loss of just 0.8%. Similarly, a shock to the skew c.r.f. of -9 together with a market c.r.f. shock of 3 yields an estimated 8.0% loss under the DFMSA approach whereas the SSA approach actually yields a gain of 0.6%.

It is also clear from Table 2 that, as designed, the portfolio has the correct directional exposure to positive moves in the parallel shift c.r.f. and negative moves in the market c.r.f. Furthermore, the portfolio also reacts positively to positive moves in the skew c.r.f. This can be explained by observing the correlations between the skew and the parallel shift c.r.f. returns as reported in Appendix B. Specifically, the skew c.r.f. return is positively correlated with the parallel shift c.r.f. ($\approx 30\%$ correlation) and has close to zero correlation to the market c.r.f. return ($\approx 2\%$). Since the portfolio is positively exposed to shocks in the parallel shift c.r.f. return it is therefore no surprise to see the portfolio is also positively exposed to positive shocks to the skew c.r.f. too. This of course is not captured by the SSA results in Table 1.

Motivated by the (partial) error decomposition in (14) from Section 5, we define

$$\mathrm{E}^{\text{abs}}_{\text{cond}} := \frac{1}{T-s}\sum_{t=s}^{T-1}\left|\Delta V_t^{\text{alt}} - \Delta V_t^{\text{ss}}\right| \qquad\qquad \mathrm{E}^{\text{abs}}_{\text{vol}} := \frac{1}{T-s}\sum_{t=s}^{T-1}\left|\Delta V_t^{\text{dfm}} - \Delta V_t^{\text{alt}}\right| \tag{27}$$

where ΔV_t^{alt} is defined in (12) and is our alternative estimated scenario P&L. We obtain ΔV_t^{alt} by setting

Table 1: Average of back-test SSA P&L $\overline{\Delta V}_{\text{ss}}$ (defined in (15)) for a portfolio that is constructed to have: (i) exposure to negative changes to the market (z_t) c.r.f. returns and positive exposure to changes in the parallel shift c.r.f. returns and (ii) to be approximately neutral (max. loss within $\pm\alpha := 2\%$ according to SSA) with respect to the pre-specified scenarios in the table. Subtable (a) displays the average SSA P&L when simultaneously stressing the market and parallel shift c.r.f. returns. Subtable (b) displays the average SSA P&L when simultaneously stressing the market and skew c.r.f. returns. All P&L numbers are in dollars per \$100 of face value of the portfolio. The portfolio is constructed anew on each day of the back-test period.

	$\overline{\Delta V}_{\text{ss}}$													
	(a) Parallel Shift							(b) Skew						
Mkt	-1.5	-1.0	-0.5	0	0.5	1.0	1.5	-9	-6	-3	0	3	6	9
-3	-0.6	0.4	0.1	-0.2	-0.6	-1.1	-1.7	0.0	-0.1	-0.2	-0.2	-0.2	-0.1	0.0
-2	1.9	1.8	1.7	1.5	1.2	0.8	0.2	1.7	1.6	1.5	1.5	1.5	1.6	1.7
-1	1.6	1.7	1.6	1.5	1.3	1.0	0.5	1.8	1.7	1.6	1.5	1.5	1.5	1.5
0	0.8	0.9	1.0	1.0	0.9	0.6	0.1	1.4	1.3	1.1	1.0	0.9	0.9	0.8
1	0.8	1.0	1.2	1.2	1.0	0.7	0.2	1.8	1.6	1.3	1.2	1.0	0.9	0.8
2	0.4	0.7	0.9	0.9	0.7	0.3	-0.3	1.8	1.5	1.2	0.9	0.6	0.4	0.2
3	-1.2	-0.8	-0.6	-0.6	-0.8	-1.2	-2.0	0.6	0.2	-0.2	-0.6	-0.9	-1.3	-1.6

Table 2: Average of back-test DFMSA P&L $\overline{\Delta V}_{\text{dfm}}$ for the same portfolio and scenarios as reported in Table 1. All P&L numbers are in dollars per \$100 of face value of the portfolio.

	$\overline{\Delta V}_{\text{dfm}}$													
	(a) Parallel Shift							(b) Skew						
Mkt	-1.5	-1.0	-0.5	0	0.5	1.0	1.5	-9	-6	-3	0	3	6	9
-3	5.1	6.3	7.1	7.9	8.8	10.2	11.4	4.0	4.9	5.2	5.2	6.9	6.7	8.0
-2	3.5	4.4	5.8	7.2	8.3	9.5	10.1	3.6	4.0	4.9	5.5	6.4	6.7	7.6
-1	0.2	1.9	2.3	4.8	5.6	7.2	7.8	1.6	1.4	2.4	3.4	4.1	5.0	5.6
0	-2.7	-1.8	-0.9	1.1	2.5	3.4	4.0	-1.1	-0.8	0.8	0.7	1.6	2.4	3.0
1	-6.2	-4.4	-3.9	-1.9	-0.6	0.9	1.5	-2.5	-1.5	-1.5	-0.5	-0.4	0.4	0.8
2	-9.1	-7.8	-6.3	-4.7	-3.8	-2.6	-2.2	-5.1	-4.3	-3.2	-3.7	-3.2	-1.7	-2.0
3	-13.9	-11.9	-10.6	-9.5	-8.5	-7.1	-6.8	-8.0	-8.0	-7.9	-6.5	-6.6	-6.3	-6.1

$\boldsymbol{\Delta}\mathbf{x}_t = \mathbf{B}\boldsymbol{\mu}_t^c$ where $\boldsymbol{\mu}_t^c$ (defined in (13)) is the expected value of the c.r.f. returns conditional on the scenario. It follows from the triangle inequality in (14) for each $t = s, \ldots, T-1$ that $\mathrm{E}^{\text{abs}} \leq \mathrm{E}^{\text{abs}}_{\text{cond}} + \mathrm{E}^{\text{abs}}_{\text{vol}}$.

Tables 3 and 4 display the average values of $\mathrm{E}^{\text{abs}}_{\text{vol}}$ and $\mathrm{E}^{\text{abs}}_{\text{cond}}$, respectively, in our back-test. It is clear from Table 3 that the error in reported P&L's that results from using the alternative ΔV_t^{alt} is relatively small and is less than 2.5% in all of considered scenarios. In contrast, the errors in Table 4 are significantly larger. These observations suggest (at least in this example), that the main source of error in the SSA approach is in setting the non-stressed factors to zero rather than their expectations conditional on the given scenario.

We can also observe from Table 4 (a) that the largest absolute errors occur when the market and parallel shift c.r.f. returns are subjected to the most extreme shocks. This indicates that setting the skew and term-structure c.r.f. returns to zero (which is how SSA would proceed) results in higher errors when the market and parallel shift c.r.f. returns are more severely stressed. Referring to the c.r.f. return correlation matrix $\boldsymbol{\rho}_{\boldsymbol{\eta}}$ that is reported in Appendix B, we see this observation can be explained by noting that the parallel shift c.r.f. return is moderately and negatively correlated with the term-structure c.r.f. ($\approx -33\%$ correlation) and is strongly correlated with the skew c.r.f. return ($\approx 55\%$ correlation). Clearly setting the term-structure and skew c.r.f. returns to zero would be highly inaccurate in this setting.

Table 3: Average back-test error E^{abs}_{vol} of the SSA P&L for the same portfolio and scenarios as in Tables 1 and 2.

	E^{abs}_{vol}													
	(a) Parallel Shift							(b) Skew						
Mkt	-1.5	-1.0	-0.5	0	0.5	1.0	1.5	-9	-6	-3	0	3	6	9
-3	1.7	1.6	1.4	1.6	1.4	1.7	1.4	2.2	1.8	2.3	1.9	1.8	1.9	1.9
-2	1.4	1.7	1.4	2.0	1.5	1.5	1.2	1.9	1.6	1.8	1.5	1.4	1.7	1.7
-1	1.6	1.5	1.8	1.3	1.6	1.4	1.4	1.6	2.1	1.6	1.5	2.0	1.6	1.4
0	2.0	1.7	1.7	1.7	1.4	1.5	1.6	1.7	1.9	2.0	1.8	1.6	1.6	1.8
1	2.2	1.9	1.8	1.7	1.6	1.4	1.2	1.7	1.6	2.0	1.9	1.4	1.6	1.5
2	2.0	2.2	2.0	1.8	1.7	1.8	1.5	1.9	1.8	1.6	1.5	1.9	1.6	1.8
3	2.3	1.8	2.0	1.9	1.7	1.7	1.6	2.1	1.8	2.1	1.8	1.8	1.7	1.6

Table 4: Average back-test error E^{abs}_{cond} of the SSA P&L for the same portfolio and scenarios as in Tables 1, 2 and 3.

	E^{abs}_{cond}													
	(a) Parallel Shift							(b) Skew						
Mkt	-1.5	-1.0	-0.5	0	0.5	1.0	1.5	-9	-6	-3	0	3	6	9
-3	4.3	5.8	7.1	8.5	10.0	11.5	12.8	4.3	4.9	5.6	6.1	6.8	7.5	8.1
-2	1.7	2.9	4.4	5.7	7.2	8.7	10.0	2.0	2.7	3.3	4.0	4.6	5.3	6.1
-1	1.4	0.5	1.5	2.9	4.4	5.8	7.3	0.6	0.6	1.3	1.9	2.6	3.4	4.2
0	4.1	2.8	1.5	0.5	1.5	3.0	4.4	2.2	1.6	0.8	0.3	0.7	1.5	2.1
1	7.0	5.7	4.4	2.9	1.6	0.5	1.4	4.3	3.5	2.8	2.0	1.2	0.7	0.7
2	9.8	8.5	7.3	5.9	4.5	3.0	1.5	6.5	5.7	5.0	4.2	3.6	2.7	2.0
3	12.6	11.3	10.0	8.8	7.4	5.9	4.4	8.8	8.0	7.2	6.5	5.8	5.1	4.4

6.3. Historical Back-Testing

While DFMSA and SSA performed on simulated paths of the ground-truth model provides some insight into their relative performance, a comparison of both approaches on actual historical scenarios would provide more concrete support. To accomplish this, we perform both SSA and DFMSA for a selection of derivative securities during days of extreme market volatility during the 2008 financial crisis. As a benchmark, we compute the true realized P&L for each of the securities during these dates. We then compare the true realized P&L to the stressed P&Ls obtained via SSA on one hand, and to the stressed P&Ls obtained via DFMSA on the other. The objective of this comparison is to provide some direct[15] evidence regarding the extent to which DFMSA provides a better picture of the risks of a security or portfolio than SSA.

To perform the historical back-test on a specific day t, we use the GARCH-embedded state-space model described earlier and estimate the corresponding parameters of the model using a window of the $s > 0$ periods up to and excluding day t. We use the same data-sets as described in Section 6.1 to estimate the model each day. The historical stress scenario is selected by choosing a subset of c.r.f.s and setting them to their *realized* values on day $t + 1$. While this is straightforward for the observable c.r.f. return, i.e. the devolatilized S&P 500 c.r.f return, it presents a difficulty if we choose any latent c.r.f. returns to stress as the realization of the latent c.r.f. returns would have to be inferred. Using estimated c.r.f. returns, however, would induce a bias in our historical back-testing procedure that favors DFMSA over SSA. For this reason, we will only[16] consider

[15]Indirect evidence is obtained via the various statistical tests that any multivariate dynamic model can be subjected to. a sample of such tests is described in Section 7.

[16]In Appendix C we discuss the issue of inferring the realization of latent c.r.f. returns in the context of a historical back-test and why that results in a bias that overstates the accuracy of DFMSA. Our results in Appendix C suggest the bias may be quite small, however.

scenarios that stress the devolatilized S&P 500 c.r.f return here and our historical back-testing procedure is outlined in Algorithm 2 below.

Algorithm 2 Historical back-testing with observable c.r.f. returns to compare SSA and DFMSA

Input: $s, t, K, \mathbf{s}$ $\triangleright$ s = # periods in window for model training
$\triangleright$ t = period to perform SA
$\triangleright$ K = # of samples used to estimate factor model-based scenario P&L
$\triangleright$ $\mathbf{s}$ = indices of observable c.r.f. returns to stress.

Estimate DFM parameters
Estimate π_{t+1} from $(\mathbf{f}^{o}_{(t-s):t}, \mathbf{\Delta x}_{(t-s):(t-1)})$ $\triangleright$ $\mathbf{f}^{o}_{(t-s):t}$ are observable
for $k \leftarrow 1$ to K **do**
  Generate $\mathbf{f}^{(k)}_{t+1} \mid (\mathcal{F}_t, \mathbf{f}_{\mathbf{s},t+1} = \mathbf{c})$ and $\boldsymbol{\epsilon}^{(k)}_{t+1}$ to obtain $\mathbf{\Delta x}^{(k)}_t$
  Compute scenario P&L $\Delta V_t(\mathbf{\Delta x}^{(k)}_t)$
end for
Compute $\Delta V^{\text{dfm}}_t := \sum_{k=1}^{K} \Delta V_t(\mathbf{\Delta x}^{(k)}_t)/K$ $\triangleright$ Estimated scenario P&L
Compute ΔV^{ss}_t $\triangleright$ SSA P&L obtained by setting $\boldsymbol{\epsilon}_{t+1}$, non-stressed common factors to $\mathbf{0}$
Compute ΔV^{act}_t using the realized value of $\mathbf{\Delta x}_t$ $\triangleright$ Actual realized P&L
Compute errors $\mathrm{E}^{\text{dfm}}_t := \left|\Delta V^{\text{dfm}}_t - \Delta V^{\text{act}}_t\right|$ and $\mathrm{E}^{\text{ss}}_t := \left|\Delta V^{\text{ss}}_t - \Delta V^{\text{act}}_t\right|$
Compute the ratio $\mathrm{E}^{\text{ratio}}_t := \mathrm{E}^{\text{dfm}}_t / \mathrm{E}^{\text{ss}}_t$

Output: $\Delta V^{\text{dfm}}_t, \Delta V^{\text{ss}}_t, \Delta V^{\text{act}}_t, \mathrm{E}^{\text{dfm}}_t, \mathrm{E}^{\text{ss}}_t$ and $\mathrm{E}^{\text{ratio}}_t$

We selected three dates to perform this historical analysis, namely September 29, 2008, when the S&P500's log-return was -9.22%, October 13, 2008 when the S&P log-return was +10.96%, and October 15, 2008 when the S&P log-return was -9.47%. For a particular date of interest, we use a window of the $s = 250$ previous trading days to fit the state-space model (21) and (22) as described in Section 6.1. We then set the scenario of interest to be the realized market[17] c.r.f. return and then proceeded with SSA and DFMSA to obtain the stressed P&Ls ΔV^{dfm}_t and ΔV^{ss}_t. Denoting the actual time t realized P&L by ΔV^{act}_t, we calculate the absolute errors of each SA approach as

$$\mathrm{E}^{\text{dfm}}_t := \left|\Delta V^{\text{dfm}}_t - \Delta V^{\text{act}}_t\right| \qquad \mathrm{E}^{\text{ss}}_t := \left|\Delta V^{\text{ss}}_t - \Delta V^{\text{act}}_t\right|. \tag{28}$$

Finally, we compute the ratio $\mathrm{E}^{\text{dfm}}_t/\mathrm{E}^{\text{ss}}_t$ which provides a measure of the performance of DFMSA compared to SSA. For example, if the ratio is equal to 1 then both approaches provide similar errors, whereas a ratio that is smaller (greater) than 1 indicates that DFMSA gave a more (less) accurate scenario P&L estimate than SSA. The results are displayed in Table 5. In all cases we see that DFMSA outperforms SSA with $\mathrm{E}^{\text{dfm}}_t/\mathrm{E}^{\text{ss}}_t$ always less than 50% and in some cases is less than 7%. While this should come as no surprise - it's well known that S&P 500 returns and implied volatility moves are strongly negatively correlated - and so it's clear that SSA would do a particularly bad job for these scenarios. That said, these results do serve to highlight the importance of embedding scenario analysis in a dynamic factor model and accounting for dependency among factors when estimating scenario P&L.

It would be interesting to see how the results in Table 5 would change if we assumed a DFM that did not include the GARCH component for the S&P log-returns. We can easily do this by again assuming the state-space model (21) and (22) but now with the first element of $\mathbf{\Delta x}_t$ in (20) equal to the S&P 500 log-return rather than its *devolatilized* log-return. The results are displayed in Table 6. There are two clear

[17] We can back-out the realization of the market c.r.f. return, defined as the innovation term in the GARCH model (17), by taking the realized S&P 500 log-return and dividing it by the σ_t computed from the estimated GARCH model via (18).

Table 5: Historical SA back-test on three dates during the financial crisis for two out-of-the-money options with maturity = 10 months and for the portfolio described in Section 6.1. On each date the scenario is set to be the realized S&P c.r.f. return. All P&L numbers are in dollars per $100 of face value. We display the resulting SSA P&L, DFMSA P&L and the actual realized P&L for each security / portfolio. We also display the ratio of the DFMSA absolute error to the SSA absolute error.

Date	S&P Log-Return	Security	ΔV_t^{ss}	$\Delta V_t^{\mathrm{dfm}}$	$\Delta V_t^{\mathrm{act}}$	$\mathrm{E}_t^{\mathrm{dfm}}/\mathrm{E}_t^{\mathrm{ss}}$
9/29/2008	-9.22%	0.90 mness, 10 m. Put	39.8	68.4	77.2	23.6%
9/29/2008	-9.22%	1.05 mness, 10 m. Call	-56.8	-38.4	-33.7	20.2%
9/29/2008	-9.22%	LP Portfolio	-0.5	37.9	62.2	38.7%
10/13/2008	+10.96%	0.90 mness, 10 m. Put	-14.7	-34.0	-42.6	30.7%
10/13/2008	+10.96%	1.05 mness, 10 m. Call	61.1	42.0	28.1	42.3%
10/13/2008	+10.96%	LP Portfolio	-1.5	-9.6	-12.4	26.3%
10/15/2008	-9.47%	0.90 mness, 10 m. Put	19.9	42.5	40.0	12.4%
10/15/2008	-9.47%	1.05 mness, 10 m. Call	-46.5	-30.9	-29.7	7.1%
10/15/2008	-9.47%	LP Portfolio	-13.7	10.6	9.1	6.8%

observations we can make. First, we see that the non-GARCH-embedded DFM model outperforms SSA. This is satisfying as it suggests that even a DFMSA based on a relatively poor dynamic model can do a better job than SSA. Second, by comparing the results in Table 5 with those in Table 6 we can see that the more sophisticated GARCH-embedded model clearly does a better job than the non-GARCH version. In fact we see that the $\mathrm{E}_t^{\mathrm{dfm}}/\mathrm{E}_t^{\mathrm{ss}}$ numbers in Table 5 are almost 10 percentage points lower on average than the corresponding numbers in Table 6.

Table 6: Everything is identical to the set-up in Table 5 except the first component of the c.r.f. return vector now represents the S&P log-returns rather than the devolatilized S&P log-returns.

Date	S&P Log-Return	Security	ΔV_t^{ss}	$\Delta V_t^{\mathrm{dfm}}$	$\Delta V_t^{\mathrm{act}}$	$\mathrm{E}_t^{\mathrm{dfm}}/\mathrm{E}_t^{\mathrm{ss}}$
9/29/2008	-9.22%	0.90 mness, 10 m. Put	39.8	69.0	77.2	21.9%
9/29/2008	-9.22%	1.05 mness, 10 m. Call	-56.8	-37.8	-33.7	17.9%
9/29/2008	-9.22%	LP Portfolio	-10.4	29.0	63.9	47.0%
10/13/2008	+10.96%	0.90 mness, 10 m. Put	-14.7	-30.5	-42.6	43.5%
10/13/2008	+10.96%	1.05 mness, 10 m. Call	61.1	44.7	28.1	50.3%
10/13/2008	+10.96%	LP Portfolio	12.3	-4.9	-15.1	37.4%
10/15/2008	-9.47%	0.90 mness, 10 m. Put	19.9	38.1	40.0	9.3%
10/15/2008	-9.47%	1.05 mness, 10 m. Call	-46.5	-31.6	-29.7	11.4%
10/15/2008	-9.47%	LP Portfolio	-10.4	2.6	10.0	36.2%

7. Statistical Evaluation of the Model in DFMSA

While not our main focus, a key aspect to implementing DFMSA in practice is the statistical evaluation of the dynamic factor model (d.f.m.) in question. We have argued that the SSA approach does not require or use a probabilistic model (see (3)) and therefore does not lend itself to any form of statistical testing. This is not true of DFMSA and in this section we briefly outline some potential approaches to the statistical validation of the underlying d.f.m. At a high level a data-set will consist of observations $(\boldsymbol{\Delta}\mathbf{x}_t, \mathbf{f}_t^o)$ for $t = 1, \ldots, T$ of the risk factor returns and observable c.r.f. returns. While most of the state-space model literature, e.g.

[34, 32], tends to focus on the estimation and implementation of these models there appears to be relatively little work on the statistical testing of these models. Some notable exceptions include [30, 33]. Because the ultimate goal of these models in our context is the accurate estimation of the daily P&L for a given portfolio (in a given scenario) we will focus here on some tests that can be applied to the one-dimensional time-series of portfolio returns.

7.1. VaR Exceptions for a Given Portfolio

Given any portfolio, by assumption we can use the $\mathbf{\Delta x}_t$'s to construct the univariate time series of the portfolio's realized P&L's, i.e. the $\Delta V_t(\mathbf{\Delta x}_t)$'s. As a first test of the d.f.m. it seems reasonable to require that, at the very least, the *realized* $\Delta V_t(\mathbf{\Delta x}_t)$'s should be consistent with the *estimated* $\Delta V_t(\mathbf{\Delta x}_t)$'s predicted by the d.f.m. A straightforward and commonly used approach for doing this is through the use of so-called Value-at-Risk (VaR) exceptions. Towards this end we recall that the time t α-VaR (for a given portfolio) is the $\mathcal{F}_t$-measurable random variable $\mathrm{VaR}_{t+1}(\alpha)$ that satisfies

$$\mathbb{P}\left(\Delta V_t(\mathbf{\Delta x}_t) < \mathrm{VaR}_{t+1}(\alpha) \mid \mathcal{F}_t\right) = 1 - \alpha$$

for any fixed $\alpha \in (0,1)$. The time t α-VaR is therefore the $(1-\alpha)$-quantile of the distribution of the portfolio P&L conditional on $\mathcal{F}_t$. We define a VaR exception as the event that the realized ΔV_t is lower than $\mathrm{VaR}_{t+1}(\alpha)$ and use $\mathbb{I}_{t+1}(\alpha)$ to denote the indicator function for such an event. Specifically, we define

$$\mathbb{I}_{t+1}(\alpha) := \begin{cases} 1, & \text{if } \Delta V_t(\mathbf{\Delta x}_t) < \mathrm{VaR}_{t+1}(\alpha) \\ 0, & \text{otherwise.} \end{cases} \tag{29}$$

It follows that $\mathbb{I}_{t+1}(\alpha)$ is a Bernoulli random variable with success probability $1-\alpha$. Since the $\{\mathbb{I}_t(\alpha)\}_t$'s are adapted to the filtration $\{\mathcal{F}_t\}_{t\geq 1}$, it can in fact be easily shown[18] that they form an i.i.d. sequence of Bernoulli random variables. This result forms the basis of several simple tests for the d.f.m. under consideration.

We begin by letting $\widehat{\mathrm{VaR}}_{t+1}(\alpha)$ be our d.f.m. estimate of $\mathrm{VaR}_{t+1}(\alpha)$ conditional on $\mathcal{F}_t$ for $t = 1, \ldots, T$. For example, in the linear-Gaussian state-space models of Section 6 and Appendix D, we can use the Kalman Filter to obtain the mean vector and covariance matrix of the distribution of $\mathbf{f}_{t+1} \mid \mathcal{F}_t$. We can then use (7) and (2), respectively, to simulate K samples from the distributions of $\mathbf{f}_{t+1} \mid \mathcal{F}_t$ and $\boldsymbol{\epsilon}_{t+1}$ and from there use (1) to obtain K samples $\Delta V_t^{(1)}, \ldots, \Delta V_t^{(K)}$ of the P&L $\Delta V_t(\mathbf{\Delta x}_t)$. We then take the $(1-\alpha)$-quantile $\widehat{\mathrm{VaR}}_{t+1}(\alpha)$ of the empirical distribution obtained from these K samples as our d.f.m's estimate of $\mathrm{VaR}_{t+1}(\alpha)$.

Binomial Distribution Tests

We can construct the sequence of empirical VaR exception indicators $\hat{\mathbb{I}}_{t+1}(\alpha)$ by substituting $\widehat{\mathrm{VaR}}_{t+1}(\alpha)$ for $\mathrm{VaR}_{t+1}(\alpha)$ in (29). Under the null hypothesis that our state-space model is correct, it follows that $\sum_{t=0}^{T-1} \hat{\mathbb{I}}_{t+1}(\alpha)$ has a $\mathrm{Binomial}(T, 1-\alpha)$ distribution. We can therefore use standard tests for the binomial to test the null hypothesis. For example, Kupiec [22] describes a two-sided binomial test with test statistic

$$Z_T = \frac{\sum_{t=1}^{T} \hat{\mathbb{I}}_t(\alpha) - T(1-\alpha)}{\sqrt{T\alpha(1-\alpha)}}. \tag{30}$$

[18] For a proof of this statement see Lemma 9.5 in [23], for example.

In particular, we then reject the null hypothesis at the κ level if $|Z_T| \geq \Phi^{-1}(1-\kappa/2)$, where $\Phi(\cdot)$ denotes the standard normal CDF.

Various other tests can also be employed. For example, under the same null hypothesis that our state-space model is correct, it follows that the time between consecutive $\text{VaR}_{t+1}(\alpha)$ exceptions are independent and geometrically distributed with success probability α. This property can be tested by approximating the geometric distribution with an exponential distribution and using a Q-Q plot or a likelihood ratio test as proposed by [8]. See also [23] for further details and additional discussion of these and other tests.

Table 7 (a) shows the results of the VaR exceptions' binomial test for the GARCH-embedded dynamic factor model, as described in Section 6, where for each day t we fit the model using observable data for the previous $s = 500$ trading days. We note that in 2008 the model results in a statistically significant high number of both 99% and 95% VaR exceptions for the majority of the assets considered. However, for the remaining years the number of VaR exceptions are in accordance with the expected number of exceptions for the most part. We suspect that adding an extreme-value-theory component to the GARCH model might improve the model sufficicently to pass these tests even in the financial crisis period of 2008. See for example Section 9.3 of [23].

For comparison, Table 7 (b) shows the results for the non-GARCH-embedded model that that we also considered in Section 6.3. This model is simply the state-space model of (22) and (7) but where the first element of $\mathbf{\Delta x}_t$ and the first c.r.f. correspond to the S&P 500 log-returns themselves. We note that the non-GARCH model performs considerably worse than the GARCH-embedded state-space model. This is not surprising as the non-GARCH model is not capable of capturing stochastic and time-varying volatility and in particular, volatility clustering.

Independence Tests

Because the $(1-\alpha)$-VaR violations form an i.i.d. sequence of Bernoulli random variables with parameter α, it follows that the time-intervals between such violations will be i.i.d. geometric, again with parameter α. We can therefore construct a test of the geometric hypothesis for any proposed d.f.m. We can use the fact that a discrete-time Bernoulli process for rare events can be approximated by a Poisson process. This in turn implies that the discrete geometric distribution for the intervals between VaR violations can be approximated by an exponential distribution. To be precise, and paraphrasing from Section 9.3 of [23], we assume the time interval $[t, t+1]$ in discrete time has length Δt in the chosen unit of continuous time. For example, if $[t, t+1]$ represents a trading day, then $\Delta t = 1$ if time is measured in days and $\Delta t = 1/250$ if time is measured in years. If the Bernoulli rare event probability is $1-\alpha$, then the approximating Poisson process has rate $\lambda = (1-\alpha)\Delta t$ and the approximating exponential distribution has mean $1/\lambda$. The exponential hypothesis can be tested using a Q-Q plot of the intervals between $(1-\alpha)$-VaR violations against the quantiles of a standard exponential distribution.

Figure 1 shows the Q-Q plots for the 95%-VaR independence test for the same securities as those examined in Table 7 for the period January 2008 through August 2013. To aid in visualizing the results, the plots are on a logarithmic scale. We observe that the S&P 500 index, as well as the 6 and 9 month put options with moneyness 0.9 generally have slightly shorter spacings between VaR exceptions than the exponential distribution benchmark. In addition, most of the options have a heavier right tail than the exponential distribution benchmark, so that the longest spacings between VaR exceptions are larger than they should be under the exponential assumption. These results are not surprising in light of the results from Table 7

Table 7: Number of 95% and 99% VaR exceptions of (a) the GARCH-embedded d.f.m, where we use the innovations z_t as a c.r.f., and (b) the non-GARCH d.f.m., where we use the log-returns of the S&P 500 index as a c.r.f. We highlight significant differences between the expected and realized number of exceptions, according to the binomial test at the 5% confidence level.

	(a) GARCH-Embedded D.F.M.							(b) Non-GARCH D.F.M.						
Year	08	09	10	11	12	13*	All	08	09	10	11	12	13*	All
	95% VaR Exceptions							95% VaR Exceptions						
Expected	13	13	13	13	13	8.5	71	13	13	13	13	13	8.5	71
S&P500 index	**27**	15	14	18	10	9	**93**	**34**	**4**	13	**22**	**4**	10	87
0.90 mness, 6m Put	**27**	6	6	**20**	10	9	78	**26**	**4**	14	**22**	**5**	**16**	87
1.05 mness, 6m Call	18	9	6	10	8	5	56	**39**	**2**	**4**	19	7	3	74
0.90 mness, 9m Put	19	11	7	14	10	6	67	**24**	**4**	12	**22**	**4**	13	79
1.05 mness, 9m Call	19	9	6	12	9	5	60	**39**	**2**	**5**	17	6	4	73
0.90 mness, 12m Put	**20**	14	11	15	8	9	77	18	**4**	10	18	**3**	4	57
1.05 mness, 12m Call	**20**	13	6	10	12	6	67	**42**	**4**	8	**22**	**5**	6	87
	99% VaR Exceptions							99% VaR Exceptions						
Expected	2.6	2.6	2.6	2.6	2.6	1.7	14.3	2.6	2.6	2.6	2.6	2.6	1.7	14.3
S&P500 index	**8**	4	**8**	**7**	5	2	**34**	**20**	0	**7**	**14**	1	3	**45**
0.90 mness, 6m Put	5	4	2	2	4	0	17	**14**	3	5	**12**	1	**7**	**42**
1.05 mness, 6m Call	**6**	4	1	3	3	0	17	**18**	2	2	**10**	1	0	**33**
0.90 mness, 9m Put	**6**	4	2	2	4	1	19	**14**	2	4	**12**	1	2	**35**
1.05 mness, 9m Call	**6**	5	1	2	3	0	17	**18**	1	2	**12**	1	1	**35**
0.90 mness, 12m Put	**7**	1	1	4	1	1	15	**10**	1	3	**9**	1	1	**25**
1.05 mness, 12m Call	**6**	3	1	3	3	3	19	**17**	1	3	**10**	1	2	**34**

*Options data was available through August 2013.

where we noted that there were a significant number of VaR exceptions in 2008. It stands to reason then that in the Q-Q plots of Figure 1, we would expect the large number of 2008 VaR exceptions to result in intervals between VaR exceptions being shorter than the exponential distribution would suggest.

7.2. Scenario VaR Exceptions

We can use the same VaR exception framework to evaluate the state-space model within the context of scenario analysis. In particular, instead of calculating the VaR from the distribution of $\Delta V_t(\boldsymbol{\Delta}\mathbf{x}_t) \mid \mathcal{F}_t$, we use the distribution of $\Delta V_t(\boldsymbol{\Delta}\mathbf{x}_t) \mid (\mathcal{F}_t, \mathbf{f}_{\mathbf{s},t+1} = \mathbf{c}_{t+1})$ for some subset $\mathbf{s}$ of the c.r.f. vector $\mathbf{f}_{t+1}$ and for some time $t+1$ scenario $\mathbf{c}_{t+1}$. In order to count the VaR exceptions, however, we must be able to obtain the realization of the P&L conditional on $\mathcal{F}_t$ and $\mathbf{f}_{\mathbf{s},t+1} = \mathbf{c}_{t+1}$. We therefore must set $\mathbf{c}_{t+1}$ to be equal to the realized value $\mathbf{f}^r_{\mathbf{s},t+1}$ of $\mathbf{f}_{\mathbf{s},t+1}$ as we did in the historical back-testing framework of Section 6.3. Moreover, we only[19] consider scenarios where only observable c.r.f. returns are stressed. For each time t we estimate $\widehat{\mathrm{VaR}}_{t+1}(\alpha) \mid (\mathcal{F}_t, \mathbf{f}_{\mathbf{s},t+1} = \mathbf{f}^r_{\mathbf{s},t+1})$ again using Monte Carlo as described in Section 7.1 but where we now sample from $\mathbf{f}_{t+1} \mid (\mathcal{F}_t, \mathbf{f}_{\mathbf{s},t+1} = \mathbf{f}^r_{\mathbf{s},t+1})$. Having estimated each scenario-conditional $\widehat{\mathrm{VaR}}_{t+1}(\alpha)$, we can compute the empirical VaR exception indicator $\hat{\mathbb{I}}_{t+1}(\alpha)$ and conduct the same tests as described in Section 7.1.

Table 8 (a) shows the results of the VaR exceptions' binomial test for the GARCH-embedded dynamic factor model. In this case the only observable c.r.f. return is the devolatilized S&P 500 log-return and so for

[19]This is because if scenarios include the stressing of latent c.r.f. returns, then it will be necessary to estimate the realized values of these c.r.f. returns which could potentially introduce significant bias into the statistical testing of the scenario VaR exceptions.

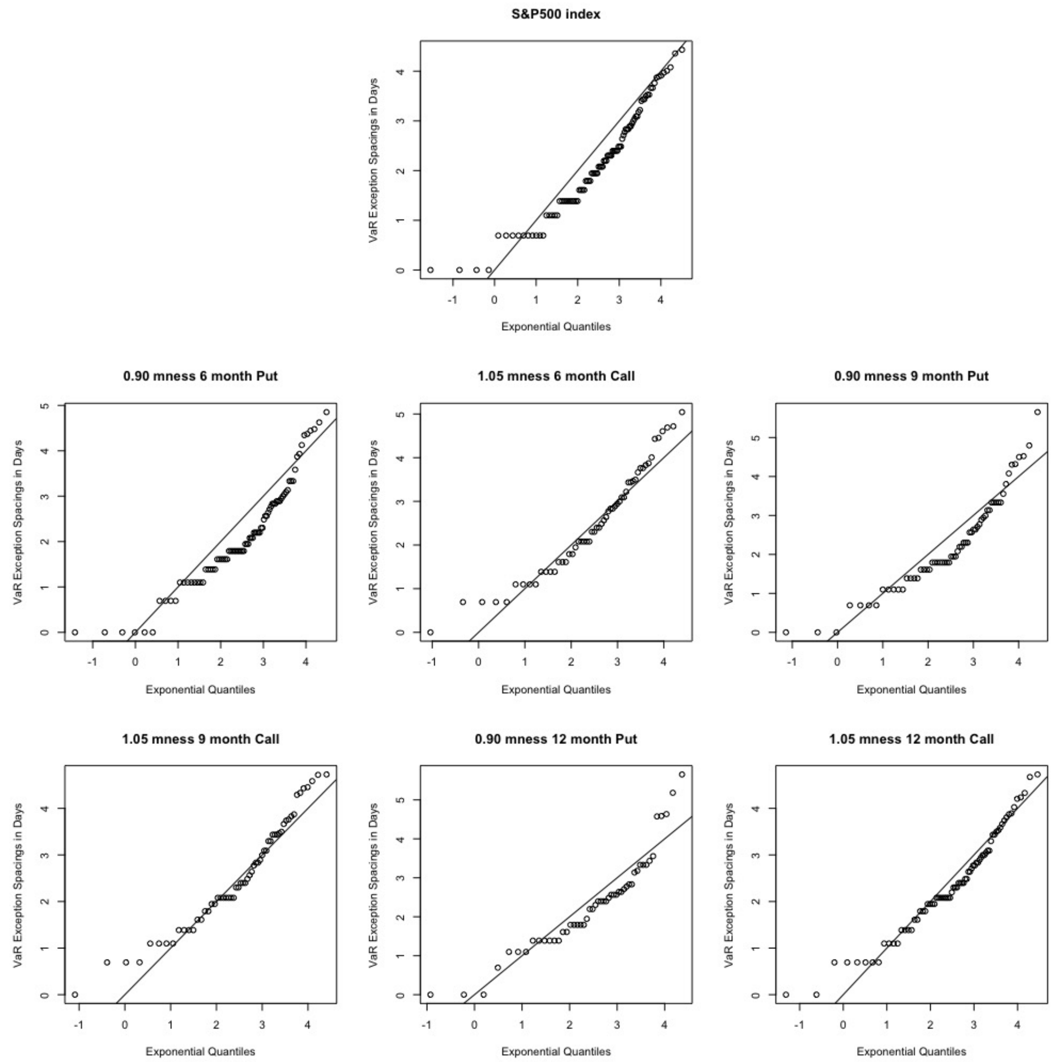

Figure 1: Independence Tests: Q-Q plots for the number of days between 95% VaR exceptions on a log scale for the S&P 500 and European call and put options on it of varying moneyness and maturity.

each day t in the back-test, the scenario in question is simply the realized devolatilized S&P 500 log-return over the period $[t, t+1]$. The results are similar to those in Table 7 (a) although one notable difference is that we can see from Table 8 (a) that occasionally there were too *few* scenario 95% VaR exceptions. Given these results were based on conditioning on the S&P 500 c.r.f. return, it suggests that perhaps a more sophisticated model for the dynamics of the implied volatility c.r.f. returns is required. For comparison we show in Table 8 (b) the results for the corresponding non-GARCH embedded model. Not surprisingly, this model performs far worse than the GARCH-embedded model.

We note that it would also be possible to perform independence tests of the scenario VaR exceptions as described in Section 7.1. We conclude this section by emphasizing the that any DFMSA should be accompanied by a statistical testing framework to validate the use of the d.f.m. under consideration. If it

Table 8: Number of 95% and 99% market c.r.f.-conditional VaR exceptions of (a) the GARCH-embedded d.f.m, where we use the innovations z_t as a c.r.f., and (b) the non-GARCH d.f.m., where we use the log-returns of the S&P 500 index as a c.r.f. We use the same selection of out-of-the-money options as in Table 7. We highlight significant differences between the expected and realized number of exceptions, according to the binomial test at the 5% confidence level.

	(a) GARCH-embedded d.f.m.							(b) Non-GARCH d.f.m.						
Year	08	09	10	11	12	13*	All	08	09	10	11	12	13*	All
	95% VaR Exceptions							95% VaR Exceptions						
Expected	13	13	13	13	13	8.5	71	13	13	13	13	13	8.5	71
0.90 mness, 6m Put	**26**	7	7	**23**	13	10	86	**27**	**5**	18	**30**	16	**16**	**112**
1.05 mness, 6m Call	**23**	14	**3**	15	7	3	65	**32**	**5**	16	**29**	14	9	**105**
0.90 mness, 9m Put	**27**	13	**5**	15	8	5	73	**30**	**5**	17	**26**	17	13	**108**
1.05 mness, 9m Call	**31**	15	**2**	13	7	**1**	69	**31**	**4**	14	**22**	15	5	**91**
0.90 mness, 12m Put	**28**	19	**2**	9	**2**	**0**	60	**25**	**3**	9	17	6	**2**	62
1.05 mness, 12m Call	**24**	17	**2**	9	**4**	**1**	57	**23**	**2**	8	14	10	3	60
	99% VaR Exceptions							99% VaR Exceptions						
Expected	2.6	2.6	2.6	2.6	2.6	1.7	14.3	2.6	2.6	2.6	2.6	2.6	1.7	14.3
0.90 mness, 6m Put	**14**	0	1	**10**	2	3	**30**	**15**	2	**11**	**14**	7	**10**	**59**
1.05 mness, 6m Call	**15**	2	0	**8**	0	0	**25**	**21**	2	**8**	**14**	5	4	**54**
0.90 mness, 9m Put	**14**	0	1	**8**	1	0	**24**	**20**	2	**10**	**11**	3	**7**	**53**
1.05 mness, 9m Call	**13**	4	0	4	0	0	21	**15**	1	**7**	**11**	5	3	**42**
0.90 mness, 12m Put	**11**	4	0	5	0	0	20	**13**	2	**6**	**7**	2	2	**32**
1.05 mness, 12m Call	**13**	5	0	3	0	0	21	**15**	0	4	**9**	2	2	**32**

*Options data was available through August 2013.

fails to pass the testing framework and this failure is deemed significant[20], then alternative models should be developed and tested.

8. Conclusions and Further Research

We have argued in this paper for the embedding of scenario analysis inside a dynamic factor model framework so that more accurate estimates of scenario P&L's can be computed and so that these estimates can be subjected to a rigorous back-testing framework. This is particularly important for derivatives portfolios where scenario analysis is often performed in a naive fashion by implicitly setting unstressed c.r.f. returns to zero in the absence of any dynamic model for the c.r.f. returns.

There are many interesting directions for future research. First, it would be particularly interesting to extend and develop the state-space modeling framework to more complex asset classes than considered in Section 6 and Appendix D. For example, we would like to perform DFMSA for portfolios consisting of options and equity positions on many stocks or portfolios of spot and option positions on the major FX currency pairs. It would also be of interest to extend these models to allow for stochastic correlation which is a well-observed feature of financial markets, particularly in times of crisis.

A second direction, which is complimentary to the first, would be the development of more sophisticated state-space models that move us beyond the linear-Gaussian framework we have considered in this paper. Such models would surely be required for modeling stochastic correlation but we suspect they would also

[20]We have in mind here the distinction between statistical and practical significance. Given that all models are "wrong", it's inevitable that even good models will eventually fail tests of statistical significance given sufficient data. It is important then to consider issues of practical significance when accepting or rejecting models.

lead to superior back-testing results than those obtained in Section 7. Specifically, while we used a GARCH model to devolatilize the S&P return component, we made no such correction for the three implied volatility c.r.f. returns and assumed they were sufficiently well modelled by the linear-Gaussian framework. Clearly, this assumption might not be justified and it would be interesting to develop a more sophisticated state-space model even for this single-stock application.

Of course the Kalman filtering framework no longer applies once we go beyond the linear-Gaussian state-space model. Nonetheless, given modern computing power this should not be too problematic as MCMC and particle filters can be employed instead. Moreover there is already a substantial literature, e.g. Chapter 12 of Tsay [34], on the use of MCMC and particle filters for financial time-series from which we could borrow. Instead of using MCMC and particle filters, it might be more convenient to simply set the unstressed latent c.r.f. returns equal to their conditional means or MAPs (maximum a posteriori) estimates. In fact we already experimented with this in Tables 3 and 4 of Section 6.2 and we saw there that there was only a small loss in accuracy when we set the the unstressed latent c.r.f. returns equal to their conditional means. It was easy to determine their conditional means in the linear-Gaussian set-up of that section since the conditional distributions were Gaussian and easy to compute via the Kalman filter. For more general state-space models, computing the means or MAPs would be more problematic. Nonetheless, some special purpose optimization algorithms have been developed for this purpose. For example, Aravkin [2, 3] computes the MAP estimates when the innovations have a multivariate t distribution. It is worth emphasizing that one major difficulty with moving beyond the linear-Gaussian framework is how the parameters of these models can be estimated and whether or not there is sufficient data available for their estimation. We used the well-known EM algorithm in our linear-Gaussian setting but there appears to be relatively little work on how to estimate model parameters in non-linear Gaussian settings.

We spent some time in Section 7 discussing some of the standard statistical tests to which we could subject our dynamic factor models. There we focussed mainly on VaR and so-called scenario VaR exceptions for fixed portfolios. Working with fixed portfolios can loosely be viewed as working with one-dimensional projections of the underlying multivariate distributions. This seems like a reasonable approach, particularly in light of the Cramér-Wold Theorem [10] which states that a multivariate distribution is uniquely determined by its one-dimensional projections. It would be interesting to consider other statistical approaches to model validation, however. In recent years, for example, there has been considerable interest in the concept of *elicitability* [18] and its use in model selection and comparison via back-testing; see for example [11, 4, 36]. It would be interesting to take an elicitability approach to model selection for the purposes of scenario analysis as proposed in this paper.

A further statistical issue of interest is the bias that arises in the context of historical back-testing when we have to infer the realized values of latent c.r.f. returns. We discussed this issue in Section 6.3 and provided some results in Appendix C that indicated that the size of the this bias may not be very large in some circumstances. It would be interesting to explore this issue more systematically and possibly with a view to estimating and therefore controlling for the bias in such historical back-tests.

Another interesting direction for future research relates to the recent work of Rebonato, e.g. [28, 29]. Rebonato has proposed the use of graphical models for scenario analysis in a context where macro-economic and systemic risk factors might be stressed. It might be interesting to embed our DFMSA approach within a graphical model (GM) framework. Consider for example a GM where the root nodes represent macro-economic factor returns and suppose we want to consider a scenario where some of these returns are set to some fixed stress levels. We can propagate this scenario through the other nodes of the GM via Monte-Carlo simulation to obtain samples of stresses to some c.r.f returns (represented by other nodes) in the GM. Each such sample could then represent a "scenario" in our DFM where we now need to simulate from the c.r.f. returns that were not included in the GM. A key assumption here is that the distribution of the non-GM c.r.f. returns is independent of all random variables in the GM conditional on the GM's c.r.f. returns.

More generally, it would be interesting to see if we could include all of the DFM's c.r.f. returns within the GM and to define the conditional distributions in the GM in such a way that the joint marginal distribution of the c.r.f. returns in the GM agrees with the joint distribution of the c.r.f. returns in the DFM. If we could do this, then we could perform scenario analysis at both the macro-level (which may be of more interest at the institution level) and c.r.f. return level (which may be of more interest at the desk / portfolio level) and to do so in a mutually consistent way. We suspect this should be possible if everything was multivariate normally distributed but it may be too challenging to do this more generally.

Finally, we mention the possibility of using DFMSA to establish margining rules for exchanges. A key task for derivative exchanges is constructing reasonable rules for establishing the margin requirements for portfolios of (derivative) securities. This is not an easy task and yet in order to remain competitive with other exchanges, it is important that exchanges not impose excessive requirements. We believe a scenario analysis approach, suitably embedded in a dynamic factor model, might be a useful approach for establishing such rules.

References

[1] R.A. Alfaro and M. Drehmann. Macro stress tests and crises: What can we learn. *BIS Quarterly Review*, December 2009.

[2] A.Y. Aravkin, J.V. Burke, and G. Pillonetto. Optimization viewpoint on kalman smoothing, with applications to robust and sparse estimation. In A.Y. Carmi, L.S. Mihaylova, and S.J. Godsill, editors, *Compressed Sensing and Sparse Filtering*, pages 237–280. Springer, Berlin, Germany, 2014.

[3] A.Y. Aravkin, J.V. Burke, and G. Pillonetto. Robust and trend following student's t kalman smoothers. *SIAM J. Control Optim.*, 52(5):2891–2916, 2014.

[4] F. Bellini and V. Bignozzi. On elicitable risk measures. *Quantitative Finance*, 15(5):725–733, 2015.

[5] G. Bonti, M. Kalkbrener, C. Lotz, and G. Stahl. Credit risk concentrations under stress. *Journal of Credit Risk*, 2:115–136, 2006.

[6] C. Borio, M. Drehmann, and K. Tsatsaronis. Technical report.

[7] T. Breuer, M. Jandacka, K. Rheinberger, and M. Summer. How to find plausible, severe, and useful stress scenarios. *International Journal of Central Banking*, 5:205–224, 2009.

[8] P.F. Christoffersen and D. Pelletier. Backtesting value-at-risk: a duration-based approach. *Journal of Financial Econometrics*, 2(1):84–108, 2004.

[9] R. Cont and J. Da Fonseca. Dynamics of implied volatility surfaces. *Quantitative Finance*, 2(1):45–60, 2002.

[10] H. Cramér and H. Wold. Some theorems on distribution functions. *Journal of the London Mathematical Society*, 11(4):290–294, 1936.

[11] M.H.A. Davis. Consistency of risk measure estimates. Preprint (arXiv: 1410.4382v1), 2014.

[12] E. Derman and M.B. Miller. *The Volatility Smile*. John Wiley & Sons, 2016.

[13] F. X. Diebold and C. Li. Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130:337–364, 2006.

[14] European Central Bank (ECB).

[15] M.D. Flood and G. Korenko. Systematic scenario selection: stress testing and the nature of uncertainty. *Quantitative Finance*, 15(1):43–59, 2015.

[16] J. Da Fonseca and K. Gottschalk. A joint analysis of the term structure of credit default swap spreads and the implied volatiltiy surface. *Journal of Futures Markets*, 33(6):494–517, 2013.

[17] P. Glasserman, C. Kang, and W. Kang. Stress scenario selection by empirical likelihood. *Quantitative Finance*, 15(1):25–41, 2015.

[18] T. Gneiting. Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762, 2011.

[19] B. Golub, D. Greenberg, and R. Ratcliffe. Market-driven scenarios: An approach for plausible scenario construction. *The Journal of Portfolio Management*, 2018.

[20] N. J. Gordon, D. J. Salmond, and A. F. Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEE Proceedings F (Radar and Signal Processing)*, 140(2):107–113, 1993.

[21] R. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering, Transactions of the ASME*, 82:35–46, 1960.

[22] P.H. Kupiec. Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 3(2):73–84, 1995.

[23] A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative risk management: concepts, techniques, and tools*. Princeton university press, 2015.

[24] S. Natenberg. *Option Pricing and Volatility: Advandec Trading Strategies and Techniques*. McGraw-Hill, 2nd edition, 1994.

[25] Basel Committee on Banking Supervision. International convergence of capital measurement and capital standards: A revised framework. Technical report, 2005.

[26] K. Popper. *The Logic of Scientific Discovery*. Routledge, 1959.

[27] M. Quagliarello. *Stress-Testing the Banking System*. Cambridge University Press, Cambridge, U.K., 2009.

[28] R. Rebonato. *Coherent Stress Testing: A Bayesian Approach to the Analysis of Financial Stress*. John Wiley & Sons, 2010.

[29] R. Rebonato. A financially justifiable and practically implementable approach to coherent stress testing. *Quantitative Finance*, 19(5):827–842, 2019.

[30] Alejandro Rodriguez and Esther Ruiz. Bootstrap prediction intervals in state-space models. *Journal of Time Series Analysis*, 30(2):167–178, 2009.

[31] L.C.G. Rogers and M.R. Tehranchi. Can the implied volatility surface move by parallel shifts? *Finance and Stochastics*, 14(2):235–248, 2010.

[32] Robert H. Shumway and David S. Stoffer. *Time Series Analysis and Its Applications: With R Examples*. Springer International Publishing, 4 edition, 2017.

[33] David S. Stoffer and Kent D. Wall. Bootstrapping state-space models: Gaussian maximum likelihood estimation and the kalman filter. *Journal of the American Statistical Association*, 86(416):1024–1033, 1991.

[34] R. S. Tsay. *Analysis of Financial Time Series*. John Wiley & Sons, 2010.

[35] E. Wan and R. Van der Merwe. The unscented kalman filter for nonlinear estimation. *Proc. Adaptive Systems for Signal Process., Commun., and Control Symp.. AS-SPCC.*, pages 153–158, 2000.

[36] J.F. Ziegel. Coherence and elicitability. *Mathematical Finance*, 26(4):901–918, 2016.

A. Portfolio Construction Via Linear Programming

We develop a simple linear programming (LP) approach to construct a portfolio according to the setting and notation introduced in Section 5.2. We assume the p.m can trade in N securities and that their daily P&L, $\Delta v_i(\mathbf{\Delta x})$, for $i = 1, \ldots, N$, depends on the vector of risk factor changes $\mathbf{\Delta x} \in \mathbb{R}^n$. The p.m. wishes to determine the portfolio weights $w_1, \ldots, w_N$ where $\sum_{i=1}^N w_i = 1$ and where w_i is the percentage of the portfolio value allocated to security i. The p.m. believes $\mathbf{f_e} = \mathbf{c}$ at the end of the next period and wishes to construct per portfolio to take advantage of this belief. The p.m. also believes and uses the DFM approach and has therefore estimated π_{t+1} as well as the parameters of the model (2) and the corresponding dynamic factor model for $\mathbf{f}_t$. She can therefore easily simulate K samples of the risk factor changes, $\mathbf{\Delta x}^{(1)}, \ldots, \mathbf{\Delta x}^{(K)}$ and the use these samples to estimate the expected P&L for each of the N securities conditional on the view, i.e. scenario, that $\mathbf{f_e} = \mathbf{c}$. We let $\Delta v_i^{\text{fm}} := \frac{1}{K}\sum_{k=1}^K \Delta v_i(\mathbf{\Delta x}^{(k)})$ denote these expected conditional P&Ls. Letting $\mathbf{w} := (w_1, \ldots, w_N)$, the p.m.'s objective function will therefore be given by

$$F(\mathbf{w}) := \sum_{i=1}^N w_i \Delta v_i^{\text{fm}} \tag{31}$$

which is her expected portfolio P&L conditional on the view $\mathbf{f_e} = \mathbf{c}$.

The p.m. must also satisfy certain constraints imposed by the risk management team. In particular, the risk management team require that the estimated scenario P&L's for L different scenarios must lie between $-\alpha\%$ and $\alpha\%$. These estimated scenario P&L's are computed using the SSA approach and involve stresses to combinations of the c.r.f.'s in $\mathbf{f_n}$. For each security $i = 1, \ldots, N$ and each scenario $l = 1, \ldots, L$, we can use the SSA approach to estimate the P&L of the i^{th} security in that scenario. If we denote this estimated P&L by $\Delta v_i^{(l)}$ then these risk constraints will result in the following linear constraints for the LP:

$$\begin{aligned} A_{l+}(\mathbf{w}) &:= \sum_{i=1}^N w_i \Delta v_i^{(l)} \leq \alpha \quad \text{for } l = 1, \ldots, L \\ A_{l-}(\mathbf{w}) &:= \sum_{i=1}^N w_i \Delta v_i^{(l)} \geq -\alpha \quad \text{for } l = 1, \ldots, L \end{aligned} \tag{32}$$

We can then combine (31) and (32) together with the constraint $\mathbf{1}^\top \mathbf{w} = 1$ to obtain the full LP that the p.m. must solve to obtain her optimal portfolio.

We note that it's easy to formulate more realistic LPs. For example, it would make sense to allow α to be scenario dependent and only limit the downside risk in the L scenarios. Similarly, we could assume the risk-management team is more sophisticated and therefore use DFMSA when estimating the scenario P&Ls. Likewise, it is easy to include constraints imposed by the p.m. rather than the risk-management team. Additional constraints on the so-called Greeks, e.g. delta, gamma, vega etc, of the overall portfolio as well as position constraints could also be imposed in the LP. Nonetheless the LP formulated above seems like a very straightforward way to highlight the problems that can arise when using SSA rather than DFMSA.

B. Ground Truth Parameters for the S&P 500 Options Portfolio of Section 6

The parameters of the GARCH model (17) and (18) from Section 6, given by

$$
\begin{aligned}
r_t &= \sigma_t z_t, && t = 1, 2, \ldots \\
\sigma_t^2 &= \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2, && t = 1, 2, \ldots
\end{aligned}
$$

are estimated using the historical time series of S&P 500 returns via the standard R library *rugarch*. The resulting estimates were $\alpha_0 = 2.06 \times e^{-6}$, $\alpha_1 = 0.0975$ and $\alpha_2 = 0.8894$.

Our state-space model (21) from Section 6 assumes the observation model

$$
\boldsymbol{\Delta}\mathbf{x}_t = \begin{bmatrix} 1 \\ \mathbf{b}^o \end{bmatrix} f_{t+1}^o + \begin{bmatrix} \mathbf{0}_3^\top \\ \mathbf{B}^u \end{bmatrix} \mathbf{f}_{t+1}^u + \begin{bmatrix} 0 \\ \boldsymbol{\epsilon}_{t+1} \end{bmatrix}
$$

where the first component of $\boldsymbol{\Delta}\mathbf{x}_t$ represents the daily log-return (in percentage points) of the devolatized S&P 500 and the remaining components represent the daily changes (in volatility points) in implied volatility for $n-1$ moneyness-maturity pairs. The factor loadings matrix $\mathbf{B}^u$ corresponding to the latent c.r.f. returns is given explicitly for each moneyness-maturity pair by (22). The parameter estimates for $\mathbf{b}^o, \mathbf{G}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ in (7) and (21) were obtained via the EM algorithm where we also imposed the constraint that $\mathbf{G}$ is diagonal. While not strictly necessary, this assumption was made to help the convergence of the EM algorithm and it implies that (i) each c.r.f. return follows a univariate AR(1) process with no exogenous covariates and (ii) the dependence in $\mathbf{f}_{t+1}$ conditional on $\mathcal{F}_t$ is induced via the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$.

The ground-truth model parameters for the observation model (21) were obtained as

$$
\mathbf{b}^o = \begin{bmatrix} -1.13 \\ -0.99 \\ -0.92 \\ -0.86 \\ -0.81 \\ -0.78 \\ -0.75 \\ \vdots \end{bmatrix}
\qquad
\mathbf{B}^u = \begin{bmatrix} 3.46 & 0.20 & -1.79 \\ 3.46 & 0.10 & -1.79 \\ 3.46 & 0.05 & -1.79 \\ 3.46 & 0.00 & -1.79 \\ 3.46 & -0.05 & -1.79 \\ 3.46 & -0.10 & -1.79 \\ 3.46 & -0.20 & -1.79 \\ \vdots & \vdots & \vdots \end{bmatrix}
\qquad
\mathbf{diag}(\Sigma_\epsilon^{1/2}) = \begin{bmatrix} 0.0151 \\ 0.0044 \\ 0.0029 \\ 0.0033 \\ 0.0033 \\ 0.0040 \\ 0.0108 \\ \vdots \end{bmatrix}
$$

where we show only[21] the rows of $\mathbf{b}^o, \mathbf{B}^u$ and $\mathbf{diag}(\Sigma_\epsilon^{1/2})$ that correspond to the first seven moneyness-maturity pairs (ξ, τ): $(0.80, 30d)$, $(0.90, 30d)$, $(0.95, 30d)$, $(1.00, 30d)$, $(1.05, 30d)$, $(1.10, 30d)$ and $(1.20, 30d)$.

The estimated parameters of the c.r.f. returns model (7) are

$$
\mathbf{G} = \begin{bmatrix} -0.0789 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & -0.1093 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & -0.3556 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & -0.1432 \end{bmatrix}
\qquad
\boldsymbol{\Sigma}_{\boldsymbol{\eta}} = \begin{bmatrix} 0.00009 & 0.00000 & 0.00001 & 0.00000 \\ 0.00000 & 0.00001 & 0.00004 & 0.00000 \\ 0.00001 & 0.00004 & 0.00057 & -0.00001 \\ 0.00000 & 0.00000 & -0.00001 & 0.00001 \end{bmatrix}
$$

where the first, second, third and fourth rows (and columns) of $\mathbf{G}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$ represent the devolatized S&P

[21] The complete model parameters are available upon request.

500, parallel shift, skew and term structure c.r.f. returns, respectively. For reference, the standard deviations and correlation matrix of the innovations $\boldsymbol{\eta}_t$'s are given by

$$\mathbf{diag}(\Sigma_\eta^{1/2}) = \begin{bmatrix} 0.0096 \\ 0.0034 \\ 0.0240 \\ 0.0020 \end{bmatrix} \qquad \boldsymbol{\rho}_{\boldsymbol{\eta}} = \begin{bmatrix} 1.0000 & 0.0266 & 0.0454 & 0.0010 \\ 0.0266 & 1.0000 & 0.5494 & -0.3322 \\ 0.0454 & 0.5494 & 1.0000 & -0.3565 \\ 0.0010 & -0.3322 & -0.3565 & 1.0000 \end{bmatrix}$$

The initial distribution π_0 was assumed to be normal with a zero mean vector and a diagonal covariance matrix with all diagonal elements set to 0.005.

C. Historical Back-Tests When the Scenario Stresses Include Latent C.R.F.s

Following on from Section 6.3, we consider here how we might perform a historical back-test when one or more of the c.r.f. returns that are stressed in a given scenario are latent. Suppose the current day is t and we wish to perform a scenario analysis for the period t to $t+1$ where $\mathbf{f}_{\mathbf{s},t+1}$ denotes the subset of c.r.f. returns that are stressed in the scenario. An obvious approach would be to first estimate these c.r.f. returns on date $t+1$ via some smoothed distribution by using the observable information in the time window $t-s$ through $t+s$. Specifically, we could compute $\hat{\mathbf{f}}_{\mathbf{s},t+1} := \mathbb{E}_t[\mathbf{f}_{\mathbf{s},t+1} \mid \mathbf{f}^o_{(t-s):(t+s)}, \boldsymbol{\Delta}\mathbf{x}_{(t-s):(t+s-1)}]$, where we recall that $\mathbf{f}^o_t$ corresponds to the observable c.r.f. returns. We then define the scenario $\mathbf{c}$ to be $\mathbf{c} = \hat{\mathbf{f}}_{\mathbf{s},t+1}$ and we can then proceed to estimate the DFMSA and SSA P&Ls for date $t+1$ in this scenario. Indeed it is straightforward to adjust Algorithm 2 to account for a scenario involving latent c.r.f. returns in this way.

The Bias That Results from Using Estimated C.R.F. Returns

Unfortunately setting $\mathbf{c} = \hat{\mathbf{f}}_{\mathbf{s},t+1}$ introduces a degree of bias into our results. Indeed, by taking the scenario $\mathbf{c}$ to be the scenario that is most consistent with the observed data, we are giving an implicit advantage to DFMSA. To see this, suppose for argument's sake that the entire vector $\mathbf{f}_{\mathbf{s},t+1}$ is latent and that its true realized (but unobserved) value was $\mathbf{f}^r_{\mathbf{s},t+1} = \mathbf{0}$ where we use the superscript r to denote the *realization* of the random vector $\mathbf{f}_{\mathbf{s},t+1}$. Suppose also that the realization of $\boldsymbol{\epsilon}_{t+1}$ in (2) was extreme. Then unless the observations are particularly informative, this is likely to result in a smoothed estimate $\hat{\mathbf{f}}_{\mathbf{s},t+1}$ that is far from $\mathbf{0}$. The estimated DFMSA P&L for this scenario $\mathbf{c} = \hat{\mathbf{f}}_{\mathbf{s},t+1}$ will therefore likely be closer to the true realized P&L than it would be if the scenario was equal to the true realization of the c.r.f. return, i.e. $\mathbf{c} = \mathbf{f}^r_{\mathbf{s},t+1} = \mathbf{0}$. This is because the true realized P&L will be more consistent with $\mathbf{f}_{\mathbf{s},t+1} = \hat{\mathbf{f}}_{\mathbf{s},t+1}$ than with $\mathbf{f}_{\mathbf{s},t+1} = \mathbf{f}^r_{\mathbf{s},t+1} = \mathbf{0}$. To be clear the bias arises not because our estimated P&L for that scenario is biased but rather because of how we choose the scenario in the first place. The more accurately we can estimate $\mathbf{f}_{\mathbf{s},t+1}$, the smaller this bias will be. In fact this is why we propose using a smoothed estimate of $\hat{\mathbf{f}}_{\mathbf{s},t+1}$ rather than a filtered estimate which may seem like a more natural thing to do. And indeed if $\mathbf{f}_{\mathbf{s},t+1}$ were fully observable then the bias will disappear which is why we had no bias problem with the historical back-testing results of Section 6.3 when our scenarios only involved the observable c.r.f. return which was the devolatilized S&P 500 log-return.

We can get a sense of the magnitude of this bias as follows. The simulation results of Section 6.2 were obtained by simulating a ground truth model and within this simulation we had access to the realized latent c.r.f. returns $\mathbf{f}^r_{\mathbf{s},t+1}$. Of course these realizations were not used to generate the estimated scenario P&L's there since they were treated as being unobserved. We can, however, use them here to estimate the aforementioned bias. In particular, we can compute the mean absolute difference (MAD) between the estimated P&L and the realized P&L for three scenarios: (i) $\mathbf{c} = \mathbf{f}^r_{\mathbf{s},t+1}$ (ii) $\mathbf{c} = \hat{\mathbf{f}}_{\mathbf{s},t+1}$ and (iii) $\mathbf{c} = \hat{\hat{\mathbf{f}}}_{\mathbf{s},t+1}$ where we use $\hat{\hat{\mathbf{f}}}_{\mathbf{s},t+1}$ to denote the *filtered* estimate of $\mathbf{f}_{\mathbf{s},t+1}$. Because the filtered estimate will be noisier than the smoothed estimate, the bias we described above should result in the MAD decreasing when we go from (i) to (iii).

Table 9 illustrates the MADs of SSA and DFMSA, performed by stressing the S&P and the parallel shifts c.r.f. returns, for a set of out-of-the-money put and call options. For each scenario analysis approach (SSA and DFMSA), the scenario $\mathbf{c}$ was set to each of the true realized scenario (benchmark), the scenario given by the smoothed estimates and the scenario given by the filtered estimates, respectively. To give a sense of the magnitude of the average P&L for each security, we included the mean absolute realized P&L over the simulated period. The main take away of Table 9 is that the MADs of scenario analysis (SA) when estimating the latent c.r.f.s (either using filtered or smoothed estimates) are smaller than the benchmark. The smoothed scenario numbers are closer than the filtered scenario numbers to the true scenario numbers which is as expected since the bias decreases with a more accurate estimation of the true realized scenario.

Table 9: MADs between the realized P&L and the scenario analysis P&L (SSA and DFMSA, performed by stressing the S&P and the parallel shifts c.r.f. returns). Scenarios are set to be either the true realized c.r.f. returns, or the smoothed or filtered estimates of the c.r.f. returns

	Realized Scen.		Smoothed Scen.		Filtered Scen.		Mean Absolute
Security	SSA	DFMSA	SSA	DFMSA	SSA	DFMSA	Realized P&L
0.90 mness, 3 m. Put	3.11	3.19	3.03	3.09	2.97	3.03	6.26
1.05 mness, 3 m. Call	2.91	2.90	2.82	2.82	2.76	2.77	4.35
0.90 mness, 6 m. Put	3.32	3.37	3.22	3.27	3.15	3.20	8.67
1.05 mness, 6 m. Call	2.34	2.31	2.27	2.24	2.22	2.19	5.39
0.90 mness, 1 y. Put	2.28	2.32	2.21	2.25	2.17	2.20	3.60
1.05 mness, 1 y. Call	1.88	1.82	1.82	1.76	1.78	1.73	2.78

It is also interesting to note that in the results of Table 9, DFMSA does not outperform SSA. This can be explained by the fact that the majority of scenarios in the simulation would have c.r.f. returns that are quite small in magnitude and often close to zero. The advantages of DFMSA over SSA in such scenarios would be very small indeed and so we are not surprised to see the two approaches perform similarly. To see that DFMSA outperforms SSA in more extreme scenarios, Table 10 again illustrates the MADs but now where we consider only those simulated periods for which the realized parallel shift c.r.f. return was greater than two standard deviations. The out-performance of the DFMSA approach is clear here. It's also interesting to note that across both Tables 9 and 10, the size of the bias seems relatively small. This suggests it may be safe to perform historical back-testing of scenarios that involve latent c.r.f. returns. But clearly, considerably more research would be required before we could be confident in proposing this.

D. An Application to a Portfolio of U.S. Treasury Securities

Here we consider a simple fixed income setting where the p.m. can invest in U.S. treasury securities of n distinct maturities $\tau_1, \ldots, \tau_n$. While the resulting portfolio is not a derivatives portfolio, we can nonetheless use it to highlight the importance of using DFMSA rather than SSA even for portfolios of non-derivatives securities. The risk factor changes for any portfolio chosen by the p.m. will then be the vector $\mathbf{\Delta x}_t \in \mathbb{R}^n$ whose i^{th} component denotes the change in yield-to-maturity from dates t to $t+1$ of the zero-coupon-bond

Table 10: MADs between the realized P&L and the scenario analysis P&L, calculated as in Table 9, but considering only those periods for which the parallel shift c.r.f. return was greater than 2 standard deviations.

Security	Realized Scen. SSA	Realized Scen. DFMSA	Smoothed Scen. SSA	Smoothed Scen. DFMSA	Filtered Scen. SSA	Filtered Scen. DFMSA	Mean Absolute Realized P&L
0.90 mness, 3 m. Put	7.87	6.82	7.44	6.41	6.99	6.08	18.55
1.05 mness, 3 m. Call	7.00	6.76	6.71	6.49	6.50	6.35	13.13
0.90 mness, 6 m. Put	4.70	4.53	4.52	4.32	4.41	4.16	13.80
1.05 mness, 6 m. Call	2.82	2.57	2.63	2.40	2.47	2.24	5.86
0.90 mness, 1 y. Put	2.87	2.66	2.67	2.48	2.52	2.34	4.15
1.05 mness, 1 y. Call	2.32	1.82	2.16	1.71	2.03	1.60	2.20

maturing at time τ_i. Our first step towards specifying a dynamic factor model is to specify the c.r.f's as in (2). A principal components analysis (PCA) of yield curve data suggests there are $m = 3$ c.r.f.'s for the U.S. yield curve and that these factors can explain anywhere from 85% to 95% of the total noise in the yield curve changes. In decreasing order of importance these c.r.f.'s drive parallel, slope and curvature changes, respectively, in the yield curve. To specify a parametric model for these c.r.f.'s we will use the model of Diebold and Li [13] and modify it to include an idiosyncratic noise term $\boldsymbol{\epsilon}_{t+1}$ as in (2). The resulting yield curve model can then be written as

$$\Delta x_t(\tau) = f_{1,t+1} + \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau}\right) f_{2,t+1} + \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) f_{3,t+1} + \epsilon_{t+1}(\tau) \tag{33}$$

where $\Delta x_t(\tau)$ corresponds to the yield curve change for maturity τ, $f_{1,t+1}, f_{2,t+1}$ and $f_{3,t+1}$ are the c.r.f. returns, and $\epsilon_{t+1}(\tau)$ is the component of $\boldsymbol{\epsilon}_{t+1}$ corresponding to maturity τ. The parameter λ is a positive scalar that can be chosen[22] to optimize the fit to the yield curve across some time window. The model (33) can be written in matrix form $\boldsymbol{\Delta}\mathbf{x}_t = \mathbf{B}\mathbf{f}_{t+1} + \boldsymbol{\epsilon}_{t+1}$ (as in (2)) with $b_{i,1} := 1$, $b_{i,2} := (1 - e^{-\lambda\tau_i})/\lambda\tau_i$ and $b_{i,3} := b_{i,2} - e^{-\lambda\tau_i}$ where $b_{i,j}$ denotes the $(i,j)^{th}$ element of $\mathbf{B}$.

It's clear that $\mathbf{b}_1$, the first column of $\mathbf{B}$, can capture parallel changes to the yield curve. For example, a value of $f_{1,t+1} = 1\%$ will result in the entire yield curve increasing by 1%. The second column $\mathbf{b}_2$ captures changes in the slope of the yield curve which are driven by $f_{2,t+1}$. We can see this in the left-hand plot of Figure 2 below where we see that the loadings are monotonically decreasing in τ. This means that if $f_{2,t+1} = 1\%$, for example, then short-term yields will increase considerably more than long-term yields, thereby reducing the slope of the yield curve. If the current yield curve happened to be upward-sloping then this would result in a flattening of the yield curve. The third column $\mathbf{b}_3$ captures changes in the curvature of the yield curve which are driven by $f_{3,t+1}$. We can see this in the right-hand plot of Figure 2 where we see that the loadings are monotonically increasing in τ for the first few years after which they are monotonically decreasing. Shocks to $f_{3,t+1}$ will therefore change the curvature of the current yield curve.

Of course the c.r.f. returns are latent and so we will use (2) together with the state-space model of (7) and the observation process (33) to complete the specification of our model. Specifically, we will assume that $\boldsymbol{\eta}_t$, $\boldsymbol{\epsilon}_t$ and $\mathbf{f}_0$ are normally distributed. Unlike our equity options application of Section 6, we make no attempt here to empirically validate this model nor do we claim that it provides a particularly good approximation to reality. Instead, this example is merely intended to provide a further demonstration of how we might

[22] Diebold and Li [13] chose a value of $\lambda = 0.7308$ for the US Treasury yield curve.

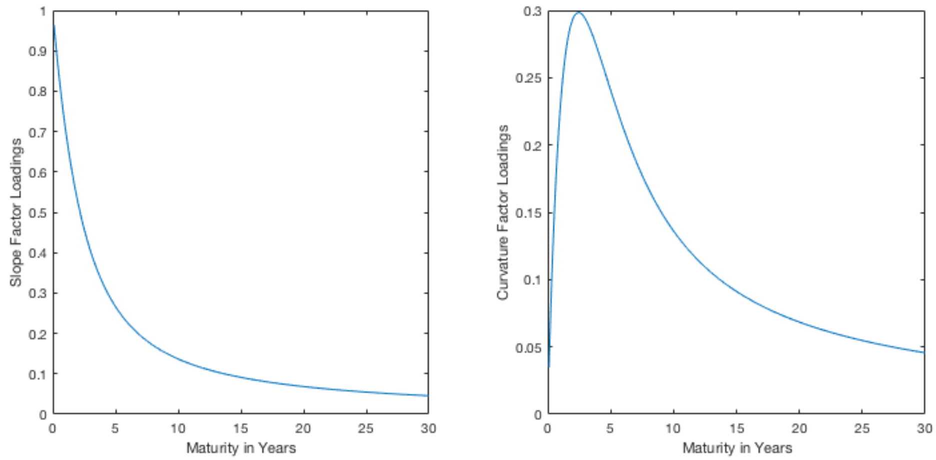

Figure 2: Factor Loadings for the Diebold factor model.

embed SSA inside a DFM and, depending on the portfolio under consideration, see that SSA can result in very misleading results.

D.1. Model Calibration and Back-Testing

In order to back-test our model we obtained US Treasury yield data from January 2008 through December 2017 for $n = 11$ maturities: 1, 3 and 6 months, and 1, 2, 3, 5, 7, 10, 20 and 30 years. We take our ground-truth model to be the model we obtain by using the EM algorithm to fit the linear Gaussian state-space model of (7) and (33) to the aforementioned yield curve data. Our yield curve model from (33) in matrix form is $\mathbf{\Delta x}_t = \mathbf{B}\mathbf{f}_{t+1} + \boldsymbol{\epsilon}_{t+1}$ where we recall $\mathbf{\Delta x}_t$ denotes the yield changes in b.p.s for the n maturities and

$$\mathbf{f}_{t+1} := \begin{bmatrix} \text{ParallelShift}_{t+1} & \text{Slope}_{t+1} & \text{Curvature}_{t+1} \end{bmatrix}^\top$$

denotes the 3×1 vector of c.r.f. returns between dates t and $t+1$. Following Diebold and Li [13] we take $\lambda = 0.7308$ which results in the loadings matrix

$$\mathbf{B} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0.97 & 0.91 & 0.84 & 0.71 & 0.53 & 0.41 & 0.27 & 0.19 & 0.14 & 0.07 & 0.05 \\ 0.03 & 0.08 & 0.14 & 0.23 & 0.29 & 0.29 & 0.24 & 0.19 & 0.14 & 0.07 & 0.05 \end{bmatrix}^\top .$$

The parameter estimates for $\Sigma_{\boldsymbol{\epsilon}}, \mathbf{G}$ and $\Sigma_{\boldsymbol{\eta}}$ in (7) were obtained from the EM algorithm and $\mathbf{G}$ was constrained to be diagonal so that each c.r.f. return follows a univariate AR(1) process with no exogenous covariates, i.e., $f_{i,t+1} = g_{i,i} f_{i,t} + \eta_{i,t+1}$ for $i = 1, 2, 3$ where $g_{i,i}$ denotes the i^{th} diagonal element of $\mathbf{G}$. As a result, the cross-sectional dependence between c.r.f.'s in $\mathbf{f}_{t+1}$ are induced exclusively via the covariance of the innovation process $\boldsymbol{\eta}_{t+1}$. The ground-truth model parameters were estimated to be

$$\mathbf{diag}(\Sigma_{\boldsymbol{\epsilon}}^{1/2}) = \begin{bmatrix} 0.0600 & 0.0312 & 0.0146 & 0.0165 & 0.0158 & 0.0109 & 0.0112 & 0.0135 & 0.0107 & 0.0056 & 0.0097 \end{bmatrix}^\top$$

$$\mathbf{G} = \begin{bmatrix} 0.0383 & 0.0000 & 0.0000 \\ 0.0000 & 0.0727 & 0.0000 \\ 0.0000 & 0.0000 & 0.0399 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_{\boldsymbol{\eta}} = \begin{bmatrix} 0.0036 & -0.0038 & -0.0002 \\ -0.0038 & 0.0066 & -0.0039 \\ -0.0002 & -0.0039 & 0.0266 \end{bmatrix}.$$

The initial distribution π_0 of the c.r.f. returns was assumed to be Gaussian with mean zero and diagonal covariance matrix with diagonal elements equal to 0.01.

For each day of our back-test we construct a portfolio using the linear programming approach described in Section 5.2 and Appendix A. The securities used to build the portfolio are zero-coupon risk-free bonds for the $n = 11$ maturities listed above as well as a risk-free cash account – the $(n+1)^{st}$ security – that each day returns 0% w.p. 1. We include a cash-account because it is realistic – p.m.s always have the option to take on zero risk by keeping their funds in cash – and it also provides a simple guarantee that there is a feasible portfolio, i.e. 100% in the cash-account, that satisfies all of the risk-constraints.

We consider a p.m. that on each day t believes $f_{1,t+1}$ will equal -12 basis points (1 b.p. = .01%) and that $f_{2,t+1}$ will equal -16 b.p.s. The p.m. is therefore always anticipating a parallel fall in the yield curve combined with an increase in it's slope. Note that the magnitude of these movements for the risk factors $\boldsymbol{\Delta}\mathbf{x}_t$ can be determined via the corresponding columns of $\mathbf{B}$ in the factor model (33). For example, if i corresponds to the 30 year maturity then $b_{i,1} = 1$ and $b_{i,2} = 0.05$ so that the resulting move in the 30-year yield is $-1 \times 0.12 - 0.05 \times 0.16 = -0.128$, i.e. a fall of 12.8 b.p.s. (This assumes the third c.r.f. return $f_{3,t+1}$ and $\boldsymbol{\epsilon}_{t+1}$ are both zero.) These anticipated movements correspond to -2 standard deviation moves in each of the first two c.r.f.s. and the p.m. wishes to construct[23] her portfolio to maximize her P&L with this view in mind.

We assume: (i) the p.m. can take on short positions so that w_i can be negative for each i and (ii) a leverage limit of 10 on each risky security so that $-10 \leq w_i \leq 10$ for each i. In addition to these constraints we assume the risk-management desk requires the p.m.'s portfolio to be "neutral" with respect to several scenarios involving joint stresses to pairwise combinations of the three c.r.f.s. They define "neutral" in such a way that the SSA P&L for the specified scenarios must be within $\pm\alpha = 3\%$ of the value of the portfolio at time t. More specifically, each scenario is given by an element of the cross-product of $\Omega_{\text{ParallelShift}} \times \Omega_{\text{Slope}}$ or $\Omega_{\text{ParallelShift}} \times \Omega_{\text{Curvature}}$ where

$$\begin{aligned} \Omega_{\text{ParallelShift}} &:= \{-24, -12, 0, 12, 24\} \\ \Omega_{\text{Slope}} &:= \{-32, -16, 0, 16, 32\} \\ \Omega_{\text{Curvature}} &:= \{-64, -32, 0, 32, 64\}. \end{aligned} \tag{34}$$

The values in $\Omega_{\text{ParallelShift}}$, Ω_{Slope} and $\Omega_{\text{Curvature}}$ were calibrated to be approximately $0, \pm 2$ and ± 4 standard deviations of the three c.r.f. returns, respectively and their units are b.p.s. Once the portfolio has been constructed we then apply SSA and DFMSA on it using the following scenarios:

[23] It's worth emphasizing that our back-tests are not at all concerned with why the p.m. has this particular view or whether or not it is ever justified. The view is simply used to construct a portfolio to which we then apply SSA and DFMSA.

(i) Simultaneous stresses to the parallel shift and slope c.r.f. returns, with shocks in the cross-product of $\Omega_{\text{ParallelShift}}$ and Ω_{Slope},

(ii) Simultaneous stresses to the parallel shift and curvature c.r.f. returns, with shocks in the cross-product of $\Omega_{\text{ParallelShift}}$ and $\Omega_{\text{Curvature}}$,

Note that the same set of scenarios are used to both construct the portfolio (via constraints on the LP) and analyze the risk of the portfolio. This of course makes sense since the constraints in the LP are driven by the scenario analysis that the risk-management desk routinely performs.

We back-tested the model using Algorithm 1 from Section 5.1 and where we used the ground-truth model to simulate data for a back-test horizon of $T = 1,000$ days. We set the training window to be of size $s = 500$ days. For each time $t \in \{s, \ldots, T-1\}$, we use the EM algorithm on the observable simulated data $\boldsymbol{\Delta}\mathbf{x}_{t-s:t-1}$ to re-estimate the model parameters $\mathbf{G}$, $\boldsymbol{\Sigma}_\eta$, $\boldsymbol{\Sigma}_\epsilon$ as well as the parameters of the normal distribution π_{t-s} governing the initial state $\mathbf{f}_{t-s}$. Once the model has been (re-)trained at time t we can use the Kalman filter to calculate the mean vector and covariance matrix of the distribution of $\mathbf{f}_t \mid \boldsymbol{\Delta}\mathbf{x}_{t-s:t-1}$. Given the c.r.f. return dynamics in (7), it then follows that π_{t+1} is the convolution of the distribution of the Gaussian random variables $\mathbf{G}\mathbf{f}_t \mid \boldsymbol{\Delta}\mathbf{x}_{t-s:t-1}$ and $\boldsymbol{\eta}_{t+1}$ and is therefore also Gaussian. Note that, even though we simulate the c.r.f. returns from the ground truth model in step 3 of Algorithm 1, these are assumed unobservable and are therefore not used by the EM algorithm to re-estimate the model parameters in step 5 of the algorithm. The SSA and DFMSA approaches are then implemented in the remaining steps of the algorithm.

D.2. Numerical Results

Tables 11 to 13 display the results of our back-test. Table 11 shows the average back-tested P&L $\overline{\Delta V}_{\text{ss}}$ as reported by the SSA approach. On each day of the back-test the portfolio was constructed in such a way that the SSA loss conditional on the given scenario would be within $\pm\alpha = 3\%$. It is therefore no surprise to see that the average-back-tested P&L numbers are also within ±3% and so this portfolio strategy *appears* to have relatively little risk. In contrast Table 12 displays the true average back-tested expected P&L $\overline{\Delta V}_{\text{dfm}}$ conditional on the given scenario. These P&L numbers were computed using the DFMSA approach and we can see from them that the portfolio is not "neutral" w.r.t. the specified scenarios. For example, when the slope c.r.f. is shocked by 32 b.p.s and the parallel c.r.f. return remains flat, the SSA approach predicts a 2.6% loss whereas the DFMSA approach predicts a 4.8% loss. We see that the differences between the two approaches can differ by up to a factor of 3. Moreover, it's possible for SSA to report a scenario loss while DFMSA reports an expected scenario profit and vice versa. We also note it's possible to obtain more extreme discrepancies between the two approaches. For example, we could have the p.m. take a more extreme view on the parallel and slope c.r.f. returns or have her take a view on the slope and curvature c.r.f. returns. Joint movements of these two c.r.f. returns are not considered in any of the scenarios and so a view on these two c.r.f. returns might allow the p.m. to better game the risk-management constraints.

Table 13 displays the mean absolute error E^{abs} (as defined in (16)) of the SSA approach. Once again we observe the large errors produced by SSA. The largest error shown is 4.1% for the scenario in which the slope and parallel c.r.f. returns are stressed to -32 bps and -24bps, respectively.

Table 11: Average of back-test SSA P&L $\overline{\Delta V}_{\text{ss}}$ (defined in (15)) for a portfolio that is constructed to have: (i) exposure to negative changes to the parallel and slope c.r.f. returns and (ii) to be approximately neutral (max. loss within $\pm\alpha := 3\%$ according to SSA) with respect to the pre-specified scenarios in the table. Subtable (a) displays the average SSA P&L when simultaneously stressing the parallel and slope c.r.f. returns. Subtable (b) displays the average SSA P&L when simultaneously stressing the parallel and the curvature c.r.f. returns. All P&L numbers are in dollars per $100 of face value of the portfolio. The portfolio is constructed anew on each day of the back-test period.

	$\overline{\Delta V}_{\text{ss}}$									
	(a) Slope (bps)					(b) Curvature (bps)				
Parallel Shift (bps)	-32	-16	0	16	32	-64	-32	0	32	64
-24	3.0	1.7	0.4	-0.9	-2.2	-2.3	-1.0	0.4	1.7	3.0
-12	2.8	1.5	0.2	-1.2	-2.5	-2.5	-1.2	0.2	1.5	2.8
0	2.6	1.3	0.0	-1.3	-2.6	-2.6	-1.3	0.0	1.3	2.6
12	2.5	1.2	-0.1	-1.4	-2.7	-2.7	-1.4	-0.1	1.2	2.5
24	2.5	1.2	-0.1	-1.4	-2.7	-2.7	-1.4	-0.1	1.2	2.4

Table 12: Average of back-test DFMSA P&L $\overline{\Delta V}_{\text{dfm}}$ for the same portfolio and scenarios as reported in Table 11. All P&L numbers are in dollars per $100 of face value of the portfolio.

	$\overline{\Delta V}_{\text{dfm}}$									
	(a) Slope (bps)					(b) Curvature (bps)				
Parallel Shift (bps)	-32	-16	0	16	32	-64	-32	0	32	64
-24	7.1	4.7	2.3	-0.1	-2.4	-5.3	-3.4	-1.7	0.0	1.7
-12	6.0	4.2	1.2	-1.3	-3.6	-4.2	-2.6	-0.9	0.9	2.6
0	5.0	2.4	0.1	-2.4	-4.8	-3.4	-1.8	0.0	1.7	3.4
12	3.7	1.5	-1.1	-3.3	-5.8	-2.5	-0.7	1.0	2.8	4.4
24	2.9	0.6	-1.9	-4.3	-6.6	-1.3	0.3	2.0	3.7	5.5

Table 13: Average back-test error E^{abs} of the SSA P&L for the same portfolio and scenarios as in Tables 11 and 12. E^{abs} is defined in (16).

	E^{abs}									
	(a) Slope (bps)					(b) Curvature (bps)				
Parallel Shift (bps)	-32	-16	0	16	32	-64	-32	0	32	64
-24	4.1	3.0	2.0	0.9	0.5	3.0	2.4	2.1	1.7	1.3
-12	3.2	2.8	1.1	0.4	1.1	1.9	1.5	1.1	0.6	0.4
0	2.3	1.1	0.4	1.1	2.2	0.8	0.6	0.3	0.6	0.9
12	1.2	0.4	1.0	2.0	3.1	0.5	0.7	1.1	1.6	2.0
24	0.5	0.7	1.8	2.9	3.9	1.4	1.7	2.1	2.5	3.0