Macroeconomic Determinants of Stock Market Returns, Volatility and Volatility Risk-Premia∗

Valentina Corradi  Walter Distaso  Antonio Mele
University of Warwick  Tanaka Business School  London School of Economics

First draft: July 22, 2005
This draft: April 21, 2008
Preliminary draft

Abstract

This paper introduces a no-arbitrage framework to assess how macroeconomic factors help explain the risk-premium agents require to bear the risk of fluctuations in stock market volatility. We develop a model in which return volatility and volatility risk-premia are stochastic and derive no-arbitrage conditions linking volatility to macroeconomic factors. We estimate the model using data related to variance swaps, which are contracts with payoffs indexed to nonparametric measures of realized volatility. We find that volatility risk-premia are strongly countercyclical, even more so than standard measures of return volatility.

Keywords: realized volatility; volatility risk-premium; macroeconomic factors; no arbitrage restrictions; concentrated simulated general method of moments, block-bootstrap.

∗We thank Yacine Aït-Sahalia for advice on the estimation of multidimensional models, and Christian Julliard, Mark Salmon and seminar participants at the University of Aarhus, the University of Warwick, the 2006 Conference on Realized Volatility (Université de Montréal) and the 2006 London-Oxford Financial Econometrics Study Group for valuable comments. Antonio Mele thanks the British EPSRC for financial support via grant EP/C522958/1. The usual disclaimer applies.
1 Introduction

Understanding the origins of stock market volatility has long been a topic of considerable interest to both policy makers and market practitioners. Policy makers are interested in the main determinants of volatility and in its spillover effects on real activity. Market practitioners are mainly interested in the direct effects time-varying volatility exerts on the pricing and hedging of plain vanilla options and more exotic derivatives. In both cases, forecasting stock market volatility constitutes a formidable challenge but also a fundamental instrument to manage the risks faced by these institutions.

Many available models use latent factors to explain the dynamics of stock market volatility. For example, in the celebrated Heston’s (1993) model, return volatility is exogenously driven by some unobservable factor correlated with the asset returns. Yet such an unobservable factor does not bear a direct economic interpretation. Moreover, the model implies, by assumption, that volatility can not be forecast by macroeconomic factors such as industrial production or inflation. This circumstance is counterfactual. Indeed, there is strong evidence that stock market volatility has a very pronounced business cycle pattern, with volatility being higher during recessions than during expansions; see, e.g., Schwert (1989a and 1989b) and Brandt and Kang (2004).

In this paper, we develop a no-arbitrage model in which stock market volatility is explicitly related to a number of macroeconomic and unobservable factors. The distinctive feature of the model is that return volatility is linked to these factors by no-arbitrage restrictions. The model is also analytically convenient: under fairly standard conditions on the dynamics of the factors and risk-aversion corrections, our model is solved in closed-form, and is amenable to empirical work.

We use the model to quantitatively assess how volatility and volatility-related risk-premia change in response to business cycle conditions. Our focus on the volatility risk-premium is related to the seminal work of Britten-Jones and Neuberger (2000), which has more recently stimulated an increasing interest in the study of the dynamics and determinants of the variance risk-premium (see, for example, Carr and Wu (2004) and Bakshi and Madan (2006)). The variance risk-premium is defined as the difference between the expectation of future stock market volatility under the true and the risk-neutral probability. It quantifies how much a representative agent is willing to pay to ensure that volatility will not raise above a given threshold. Thus, it is a very intuitive and general measure of risk-aversion. Previous important work by Bollerslev and Zhou (2005) and Bollerslev, Gibson and Zhou (2004) has analyzed how this variance risk-premium is related to a number of macroeconomic factors. The authors regressed semi-parametric measures of the variance risk-premium on these factors. In this paper, we make a step further and make the volatility risk-premium be endogenously determined within our no-arbitrage model. The resulting relation between macroeconomic factors and risk-premia is richer than in the previous
papers as it explicitly accounts for the necessary no-arbitrage relations that link asset prices and, hence, return volatility, to macroeconomic factors.

In recent years, there has been an important surge of interest in general equilibrium (GE, henceforth) models linking aggregate stock market volatility to variations in the key factors tracking the state of the economy (see, for example, Campbell and Cochrane (1999), Bansal and Yaron (2004), Mele (2007), and Tauchen (2005)). These GE models are important as they highlight the main economic mechanisms through which markets, preferences and technology affect the equilibrium price and, hence, return volatility. At the same time, we do not observe the emergence of a well accepted paradigm. Rather, a variety of GE models aim to explain the stylized features of aggregate stock market fluctuations (see, for example, Campbell (2003) and Mehra and Prescott (2003) for two views on these issues). In this paper, we do not develop a fully articulated GE model. In our framework, cross-equations restrictions arise through the weaker requirement of absence of arbitrage opportunities. This makes our approach considerably more flexible than it would be under a fully articulated GE discipline. In this respect, our approach is closer in spirit to the “no-arbitrage” vector autoregressions introduced in the term-structure literature by Ang and Piazzesi (2003) and Ang, Piazzesi and Wei (2005). Similarly as in these papers, we specify an analytically convenient pricing kernel affected by some macroeconomic factors, but do not directly relate these to markets, preferences and technology.

Our model works quite simply. We start with exogenously specifying the joint dynamics of both macroeconomic and latent factors. Then, we assume that dividends and risk-premia are essentially affine functions of the factors, along the lines of Duffee (2002). We show that the resulting no-arbitrage stock price is affine in the factors. Our model is also related to previous approaches in the literature. For example, Bekaert and Grenadier (2001) and Ang and Liu (2004) formulated discrete-time models in which the key pricing factors are exogenously given. Furthermore, Mamaysky (2002) derived a continuous-time model based on an exogenous specification of the price-dividend ratio. There are important differences between these models and ours. First, our model is in continuous-time and thus avoids theoretical inconsistencies arising in the discrete time setting considered by Bekaert and Grenadier (2001). Second, a continuous-time setting is particularly appealing given our objective to estimate volatility and volatility risk-premia through measures of realized volatility. Third, Ang and Liu (2004) consider a discrete-time setting in which expected returns are exogenous to their model; in our model, expected returns are endogenous. Finally, our model works differently from Mamaysky’s because it endogenously determines the price-dividend ratio.

Estimating our model is challenging. In our model, volatility is endogenous, which makes parameters’ identification a quite delicate issue. The main difficulty we face is that return volatility arises by a rational price formation process. Therefore, all the factors affecting the aggregate stock market also affect stock market volatility. In the standard stochastic volatility models
such as that in Heston (1993), volatility is driven by factors which are not necessarily the same as those affecting the stock price - volatility is exogenous in these models. In particular, our model predicts that return volatility can be understood as the outcome of two forces which we need to tell apart from data: (i) the market participants’ risk-aversion, and (ii) the dynamics of the fundamentals. Thus, the advantage of our model (to generate, endogenously, stock market volatility) also brings an identification issue. We address identifiability by exploiting derivative price data, related to variance swaps. The variance swap rate is, theoretically, the risk-adjusted expectation of the future integrated volatility within one month, and is published daily by the CBOE since 2003 as the new VIX index. (The CBOE has re-calculated the new VIX index back to 1990.) These data allow us to identify the model.

We implement a two-stage estimation procedure. In the first step, we use data on a broad stock market index and two macroeconomic factors, inflation and industrial production, and estimate all the parameters, taking the parameters related to risk-premia adjustments as given. We implement this step by matching moments related to ex-post stock market returns, realized return volatility and the two macroeconomic factors. In the second step, we use data on the new VIX index, and the two macroeconomic factors, to estimate the risk-premia parameters. In this second step, we exploit the general ideas underlying the “realized volatility” literature to implement consistent estimators of the VIX index (see, e.g., Barndorff-Nielsen and Shephard (2007) for a survey on realized volatility). Note, the two-stage estimation procedure entails parameter estimation error. To implement an efficient estimator, then, we rely on the block bootstrap of the entire procedure.

The remainder of the paper is organized as follows. In Section 2 we develop a no-arbitrage model for the stock price, return volatility and the variance risk-premia. Section 3 illustrates the estimation strategy. Section 4 presents our empirical results, and the appendix provides technical details omitted from the main text.

2 The model

2.1 The macroeconomic environment

We assume that a number of factors affect the development of aggregate macroeconomic variables. We assume these factors form a vector-valued process \( y(t) \), solution to a \( n \)-dimensional affine diffusion,

\[
dy(t) = \kappa (\mu - y(t)) \, dt + \Sigma V(y(t)) \, dW(t),
\]

where \( W(t) \) is a \( d \)-dimensional Brownian motion \((n \leq d)\), \( \Sigma \) is a full rank \( n \times d \) matrix, and \( V \) is a full rank \( d \times d \) diagonal matrix with elements,

\[
V(y)_{(ii)} = \sqrt{\alpha_i + \beta_i^T y}, \quad i = 1, \ldots, d,
\]
for some scalars $\alpha_i$ and vectors $\beta_i$. Appendix A reviews sufficient conditions that are known to ensure that Eq. (1) has a strong solution with $V(y(t))_{ii} > 0$ almost surely for all $t$.

While we do not necessarily observe every single component of $y(t)$, we do observe discretely sampled paths of macroeconomic variables such as industrial production, unemployment or inflation. Let $\{M_j(t)\}_{t=1, 2, \ldots}$ be the discretely sampled path of the macroeconomic variable $M_j(t)$ where, for example, $M_j(t)$ can be the industrial production index available at time $t$, and $j = 1, \ldots, N_M$, where $N_M$ is the number of observed macroeconomic factors.

We assume, without loss of generality, that these observed macroeconomic factors are strictly positive, and that they are related to the state vector process in Eq. (1) by:

$$\log \left( \frac{M_j(t)}{M_j(t-12)} \right) = \varphi_j(y(t)), \quad j = 1, \ldots, N_M,$$

where the collection of functions $\{\varphi_j\}_{j}$ determines how the factors dynamics impinge upon the evolution of the overall macroeconomic conditions. We now turn to model asset prices.

### 2.2 Risk-premia and stock market volatility

We assume that asset prices are related to the vector of factors $y(t)$ in Eq. (1), and that some of these factors affect the development of macroeconomic conditions, through Eq. (2). We assume that asset prices respond passively to movements in the factors affecting macroeconomic conditions. In other words, and for analytical convenience, we are ruling out that asset prices can feed back the real economy, although we acknowledge that financial frictions can make financial markets and the macroeconomy intimately related, as in the financial accelerator hypothesis reviewed by Bernanke, Gertler and Gilchrist (1999).

Formally, we assume that there exists a rational pricing function $s(y(t))$ such that the real stock price at time $t$, $s(t)$ say, is $s(t) \equiv s(y(t))$. We let this price function be twice continuously differentiable in $y$. (Given the assumptions and conditions we give below, this differentiability condition holds in our model.) By Itô’s lemma, $s(t)$ satisfies,

$$\frac{ds(t)}{s(t)} = m(y(t), s(t)) \, dt + \frac{s_y(y(t)) \, \Sigma \, V(y(t))}{s(y(t))} \, dW(t),$$

where $s_y(y) = \begin{bmatrix} \partial \frac{\partial s}{\partial y_1} \, y, & \cdots, & \partial \frac{\partial s}{\partial y_n} \, y \end{bmatrix}^\top$ and $m$ is a function we shall determine below by no-arbitrage conditions. By Eq. (3), the instantaneous return volatility is

$$\sigma(t)^2 = \left\| \frac{s_y(y(t)) \, \Sigma \, V(y(t))}{s(y(t))} \right\|^2.$$

Next, we model the pricing kernel in the economy. The Radon-Nikodym derivative of $Q$, the equivalent martingale measure, with respect to $P$ on $\mathcal{F}(T)$ is,

$$\xi(T) \equiv \frac{dQ}{dP} = \exp \left( - \int_0^T \Lambda(\tau)^\top \, dW(\tau) - \frac{1}{2} \int_0^T \| \Lambda(\tau) \|^2 \, d\tau \right),$$
for some adapted $\Lambda(t)$, the risk-premium process. We assume that each component of the risk-premium process $\Lambda^i(t)$ satisfies,

$$\Lambda^i(t) = \Lambda^i(y(t)), \quad i = 1, \cdots, d,$$

for some function $\Lambda^i$. We also assume that the safe asset is elastically supplied such that the short-term rate $r$ (say) is constant. This assumption can be replaced with a weaker condition that the short-term rate is an affine function of the underlying state vector. This assumption would lead to the same affine pricing function in Proposition 1 below, but statistical inference for the resulting model would be hindered. Moreover, interest rate volatility appears to play a limited role in the main applications we consider in this paper.

Under the equivalent martingale measure, the stock price is solution to,

$$\frac{ds(t)}{s(t)} = \left( r - \delta(y(t)) \right) dt + \frac{s_y(y(t))}{s(y(t))} \sum V(y(t)) d\hat{W}(t), \quad (5)$$

where $\delta(y)$ is the instantaneous dividend rate, and $\hat{W}$ is a $Q$-Brownian motion.

### 2.3 No-arbitrage restrictions

There is obviously no freedom in modeling risk-premia and stochastic volatility separately. Given a dividend process, volatility is uniquely determined, once we specify the risk-premia. Consider, then, the following “essentially affine” specification for the dynamics of the factors in Eq. (1).

Let $V^-(y)$ be a $d \times d$ diagonal matrix with elements

$$V^-(y)_{(ii)} = \begin{cases} \frac{1}{V(y)_{(ii)}} & \text{if } \Pr\{V(y(t))_{(ii)} > 0 \text{ all } t\} = 1 \\ 0 & \text{otherwise} \end{cases}$$

and set,

$$\Lambda(y) = V(y) \lambda_1 + V^-(y) \lambda_2 y, \quad (6)$$

for some $d$-dimensional vector $\lambda_1$ and some $d \times n$ matrix $\lambda_2$. The functional form for $\Lambda$ is the same as in the specification suggested by Duffie (2002) in the term-structure literature. If the matrix $\lambda_2 = 0_{d \times n}$, then, $\Lambda$ collapses to the standard “completely affine” specification introduced by Duffie and Kan (1996), in which the risk-premia $\Lambda$ are tied up to the volatility of the fundamentals, $V(y)$. While it is reasonable to assume that risk-premia are related to the volatility of fundamentals, the specification in Eq. (6) is more general, as it allows risk-premia to be related to the level of the fundamentals, through the additional term $\lambda_2 y$.

Finally, we determine the no-arbitrage stock price. Under regularity conditions developed in
the appendix,\textsuperscript{1} and assuming no-bubbles, Eq. (5) implies that the stock price is,

\[ s(y) = \mathbb{E} \left[ \int_0^{\infty} e^{-rt} \delta(y(t)) \, dt \right], \tag{7} \]

where \( \mathbb{E} \) is the expectation taken under the equivalent martingale measure. We are only left with specifying how the instantaneous dividend process relates to the state vector \( y \). As it turns out, the previous assumption on the pricing kernel and the assumption that \( \delta(\cdot) \) is affine in \( y \) implies that the stock price is also affine in \( y \). Precisely, let

\[ \delta(y) = \delta_0 + \delta^T y, \tag{8} \]

for some scalar \( \delta_0 \) and some vector \( \delta \). We have:

**Proposition 1.** Let the risk-premia and the instantaneous dividend rate be as in Eqs. (6) and (8). Then, (i) eq. (7) holds, and (ii) the rational stock function \( s(y) \) is linear in the state vector \( y \), viz

\[ s(y) = \delta_0 + \delta^T (D + r I_{n \times n})^{-1} c + \delta^T (D + r I_{n \times n})^{-1} y, \tag{9} \]

where

\[ c = \kappa \mu - \Sigma \left( \begin{array}{c} \alpha_1 \lambda_{1(1)} \\ \vdots \\ \alpha_d \lambda_{1(d)} \end{array} \right)^T, \]

\[ D = \kappa + \Sigma \left[ \begin{array}{c} \lambda_{1(1)} \beta_1^T \\ \vdots \\ \lambda_{1(d)} \beta_d^T \end{array} \right]^T + I^-, \]

\( I^- \) is a \( d \times d \) diagonal matrix with elements \( I^-_{(ii)} = 1 \) if \( \Pr\{ V(y(t))_{(ii)} > 0 \text{ all } t \} = 1 \) and 0 otherwise; and, finally \( \{\lambda_{1(j)}\}_{j=1}^d \) are the components of \( \lambda_1 \).

Proposition 1 allows us to describe what this model predicts in terms of no-arbitrage restrictions between stochastic volatility and risk-premia. In particular, use Eq. (9) to compute volatility through Eq. (4). We obtain,

\[ \sigma(t)^2 = \left\| \delta^T (D + r I_{n \times n})^{-1} \Sigma V(y(t)) \right\|^2 \left[ \frac{\delta_0 + \delta^T (D + r I_{n \times n})^{-1} c}{r} + \delta^T (D + r I_{n \times n})^{-1} y(t) \right]^{-2}. \tag{10} \]

This formula makes clear why our approach is distinct from that in the standard stochastic volatility literature. In this literature, the asset price and, hence, its volatility, is taken as given,

\textsuperscript{1}These conditions relate to the volatility term \( s_y(y)^T \Sigma V(y) \) in Eq. (3). This term must satisfy integrability conditions which make the Itô’s integral in the representation of the stock price a martingale.
and volatility and volatility risk-premia are modeled independently of each other. For example, the celebrated Heston’s (1993) model assumes that the stock price is solution to,

\[
\begin{align*}
\frac{ds(t)}{s(t)} &= m(t) \, dt + v(t) \, dW_1(t) \\
\text{and } \quad dv(t)^2 &= \kappa (\mu - v(t)^2) \, dt + \sigma v(t) \left( \rho dW_1(t) + \sqrt{1-\rho^2} dW_2(t) \right)
\end{align*}
\]

for some adapted process \( m(t) \) and some constants \( \kappa, \mu, \sigma, \rho \). In this model, the volatility risk-premium is specified separately from the volatility process. Many empirical studies have followed the lead of this model (e.g., Chernov and Ghysels (2000)). Moreover, a recent focus in this empirical literature is to examine how the risk-compensation for stochastic volatility is related to the business cycle (e.g., Bollerslev, Gibson and Zhou (2005)). While the empirical results in these papers are very important, the Heston’s model does not predict that there is any relation between stochastic volatility, volatility risk-premia and the business cycle.

Our model works differently because it places restrictions directly on the asset price process, through our assumptions about the fundamentals of the economy, i.e. the dividend process in Eq. (8) and the risk-premia in Eq. (6). In our model, it is the asset price process that determines, endogenously, the volatility dynamics. For this reason, the model predicts that return volatility embeds information about risk-corrections that agents require to invest in the stock market, as Eq. (10) makes clear. We shall make use of this observation in the empirical part of the paper. We now turn to describe which measure of return volatility measure we shall use to proceed with such a critical step of the paper.

### 2.4 Arrow-Debreu adjusted volatility

In September 2003, the Chicago Board Option Exchange (CBOE) changed its volatility index VIX to approximate the variance swap rate of the S&P 500 index return.\(^2\) The new index reflects recent advances into the option pricing literature. Given an asset price process \( s(t) \) that is continuous in time (as the asset price of our model in Eq. (9)), and all available information \( \mathbb{F}(t) \) at time \( t \), define the integrated return variance on a given interval \( [t, T] \) as,

\[
IV_{t,T} = \int_t^T \left( \frac{d}{d\tau} \text{var} \left[ \log s(\tau) \mid \mathbb{F}(u) \right] \right) \, du.
\]

The new VIX index relies on the work of Bakshi and Madan (2000), Britten-Jones and Neuberger (2000), and Carr and Madan (2001), who showed that the risk-neutral probability expectation

\(^2\)If the interest rate is zero, then, in the absence of arbitrage opportunities, the variance swap rate is simply the expectation of the future integrated return volatility under the risk-neutral probability, as defined in Eq. (12) below.
of the future integrated variance is a functional of put and call options written on the asset:

$$
\mathbb{E}[IV_{t,T}|\mathcal{F}(t)] = 2 \left[ \int_0^{F(t)} \frac{P(t,T,K)}{u(t,T)} \frac{1}{K^2} dK + \int_{F(t)}^{\infty} \frac{C(t,T,K)}{u(t,T)} \frac{1}{K^2} dK \right],
$$

(13)

where $F(t) = u(t,T)s(t)$ is the forward price, $C(t,T,K)$ and $P(t,T,K)$ are the prices as of time $t$ of a call and a put option expiring at $T$ and struck at $K$, and $u(t,T)$ is the price as of time $t$ of a pure discount bond expiring at $T$.

Eq. (13) is helpful because it delivers a nonparametric method to compute the risk-neutral expectation of the integrated variance. Our model predicts that the risk-neutral expectation of the integrated variance is:

$$
\mathbb{E}[IV_{t,T}|\mathcal{F}(t)] = \int_t^T \mathbb{E}[(\sigma(u))^2 | \mathcal{F}(t)] du,
$$

where $\mathcal{F}(t)$ is the filtration generated by the multidimensional Brownian motion in Eq. (1), and $(\sigma(t))^2$ is given in Eq. (10). It is a fundamental objective of this paper to estimate our model so that it predicts a theoretical pattern of the VIX index that matches that computed through Eq. (13).

Note that as a by product, we will be able to trace how the volatility risk-premium, defined as,

$$
VRP(t) = \sqrt{\frac{1}{T-t} \left( \sqrt{E[IV_{t,T}|\mathcal{F}(t)]} - \sqrt{E[IV_{t,T}|\mathcal{F}(t)]} \right)} ,
$$

(14)

changes with changes in the factors $y(t)$ in Eq. (1).

2.5 The leading model

We formulate a few specific assumptions to make the model amenable to empirical work. First, we assume that two macroeconomic aggregates, inflation and industrial production growth, are the only observable factors (say $y_1$ and $y_2$) affecting the stock market development. We define these factors as follows:

$$
\log \left( \frac{M_j(t)}{M_j(t-12)} \right) = \log y_j(t), \quad j = 1, 2,
$$

where $M_1(t)$ is the consumer price index as of month $t$ and $M_2(t)$ is the industrial production as of month $t$. (Data for such macroeconomic aggregates are typically available at a monthly frequency.) Hence, in terms of Eq. (2), the functions $\phi_j(y) = \log y_j$.

Second, we assume that a third unobservable factor $y_3$ affects the stock price, but not the two macroeconomic aggregates $M_1$ and $M_2$. Third, we consider a model in which the two macroeconomic factors $y_1$ and $y_2$ do not affect the unobservable factor $y_3$, although we allow for simultaneous feedback effects between inflation and industrial production growth. Therefore, we set, in
Eq. (1),
\[
\kappa = \begin{bmatrix}
\kappa_1 & \bar{\kappa}_1 & 0 \\
\kappa_2 & \kappa_2 & 0 \\
0 & 0 & \kappa_3 \\
\end{bmatrix},
\]
where \(\kappa_1\) and \(\kappa_2\) are the speed of adjustment of inflation and industrial production growth towards their long-run means \(\mu_1\) and \(\mu_2\), in the absence of feedbacks, and \(\bar{\kappa}_1\) and \(\bar{\kappa}_2\) are the feedback parameters. Moreover, we take \(\Sigma = I_{3 \times 3}\) and the vectors \(\beta_{j}\) so as to make \(y_j\) solution to,
\[
dy_j(t) = \left[\kappa_j (\mu_j - y_j(t)) + \bar{\kappa}_j (\bar{\mu}_j - \bar{y}_j(t))\right]dt + \sqrt{\alpha_j + \beta_j y_j(t)}dW_j(t), \quad j = 1, 2, 3, \tag{15}
\]
where, for brevity, we have set \(\bar{\mu}_1 \equiv \mu_2, \bar{y}_1(t) \equiv y_2(t), \bar{\mu}_3 \equiv \bar{\mu}_2, \bar{y}_2(t) = y_1(t), \bar{\mu}_3 \equiv \bar{\mu}_2, \bar{y}_3(t) \equiv 0\) and, finally, \(\beta_j \equiv \beta_{jj}\). We assume that \(\Pr\{V(y(t))_{1i} > 0 \text{ all } t\} = 1\), which it does under the conditions reviewed in Appendix A.

We assume that the risk-premium process \(\Lambda\) satisfies the “essentially affine” specification in Eq. (6), where we take the matrix \(\Lambda_2\) to be diagonal with diagonal elements equal to \(\lambda_{2(j)} \equiv \lambda_{2(jj)}\), \(j = 1, 2, 3\). The implication is that the total risk-premia process defined as,
\[
\pi(y) \equiv \Sigma V(y) \Lambda(y) = \begin{pmatrix}
\alpha_1 \lambda_{1(1)} + (\beta_1 \lambda_{1(1)} + \lambda_{2(1)}) y_1 \\
\alpha_2 \lambda_{1(2)} + (\beta_2 \lambda_{1(2)} + \lambda_{2(2)}) y_2 \\
\alpha_3 \lambda_{1(3)} + (\beta_3 \lambda_{1(3)} + \lambda_{2(3)}) y_3
\end{pmatrix}, \tag{16}
\]
depends on the factor \(y_j\) not only through the channel of the volatility of these factors (i.e. \(\lambda_{2(j)}\)), but also through the additional risk-premia parameters \(\lambda_{2(j)}\).

Finally, the instantaneous dividend process \(\delta(t)\) in (8) satisfies,
\[
\delta(y) = \delta_0 + \delta_1 y_1 + \delta_2 y_2 + \delta_3 y_3. \tag{17}
\]
Under these conditions, the asset price in Proposition 1 is given by,
\[
s(y) = s_0 + \sum_{j=1}^{3} s_j y_j, \tag{18}
\]
where
\[
s_0 = \frac{1}{r} \left[\delta_0 + \sum_{j=1}^{3} s_j \left(\kappa_j \mu_j + \bar{\kappa}_j \bar{\mu}_j - \alpha_j \lambda_{1(j)}\right)\right], \tag{19}
\]
\[
s_j = \frac{\delta_j (r + \kappa_j - \bar{\kappa}_j + \beta_j \lambda_{1(j)} + \lambda_{2(j)})}{\prod_{h=1}^{3} (r + \kappa_h + \beta_h \lambda_{1(h)} + \lambda_{2(h)}) - \kappa_1 \kappa_2 (1 - \mathbb{I}_{(j=3)})}, \quad j = 1, 2, 3, \tag{20}
\]
and where \(\bar{\kappa}_j\) and \(\bar{\mu}_j\) as in Eq. (15) and, finally, \(\mathbb{I}_{(j=3)}\) equals 1 if \(j = 3\) and zero otherwise.
Note, then, an important feature of the model. The parameters \( \lambda_{(1)i} \) and \( \lambda_{(2)i} \) cannot be identified. Intuitively, the parameters \( \lambda_{(1)i} \) and \( \lambda_{(2)i} \) determine how sensitive the total risk-premium in Eq. (16) is to changes in the state process \( y \). Instead, the parameters \( \delta_i \) determine how sensitive the dividend process in Eq. (17) is to changes in \( y \). Two price processes might be made observationally equivalent through judicious choices of the risk-compensation required to bear the asset or the payoff process promised by this asset (the dividend). The next section explains how to exploit the Arrow-Debreu adjusted volatility introduced in Section 2.4 to identify these parameters.

3 Statistical inference

Let \( \lambda = \left( \left( \lambda_{1(j)}, \lambda_{2(j)} \right)_{j=1,2,3} \right) \) and \( \theta = \left( \delta_0, (\kappa_j, \mu_j, \alpha_j, \beta_j)_{j=1,2,3}, (\tilde{\kappa}_j)_{j=1,2} \right) \). Our estimation strategy relies on a two-step procedure, which we call Concentrated Simulated General Method of Moments (C-SGMM, in the sequel). In the first step, we treat \( \lambda \) as a vector of unidentified nuisance parameters, and estimate \( \theta \) by maximum likelihood, for any given value of \( \lambda \). In the second step, we use the Arrow-Debreu adjusted volatility in Section 2.4 and calibrate \( \lambda \) to reconcile the model’s predictions with the data.

3.1 Estimation of \( \theta \), for given \( \lambda \)

Let \( y_t = (y_{1,t}, y_{2,t}, y_{3,t}) \) denote the skeleton of \( y(t) \) in Eq. (2), sampled at \( t = 1, \cdots, T \), and set \( x_t = (y_{1,t}, y_{2,t}) \). Although the conditional density of \( x_{t+1} \) given \( x_t \) is unknown in closed-form,\(^3\) we may rely on a variety of approaches to approximate this density. In this paper, we use A¨ıt-Sahalia’s (2008) small-time expansions to obtain parameter estimates for the observable variables, i.e. estimates of \( \theta_O = \left( \left( (\kappa_j, \mu_j, \alpha_j, \beta_j)_{j=1,2}, (\tilde{\kappa}_j)_{j=1,2} \right) \right) \). The maximum likelihood estimator \( \hat{\theta}_{O,T} \) is,

\[
\hat{\theta}_{O,T} = \arg \max_{\theta_O \in \Theta} l_T (\theta_O)
\]

where,

\[
l_T (\theta_O) = \frac{1}{T} \sum_{t=1}^{T} l_t (x_{t+1}|x_t; \theta_O); \quad l_t (x_{t+1}|x_t; \theta_O) = \log \varphi_K (x_{t+1}|x_t; \theta_O), \quad (21)
\]

and \( \varphi_K (x_t|x_{t-1}; \theta_O) \) is the conditional density obtained through the A¨ıt-Sahalia’s expansion up to the \( K \)-th term.

To estimate the parameters \( \left( \delta_0, (\delta_j)_{j=1,2,3}, \kappa_3, \mu_3, \alpha_3, \beta_3 \right) \) related to the unobservable processes \( \delta (y) \) (the dividend) and \( y_3 \) (the third unobservable factor), we proceed as follows. First, we note by Eqs. (19)-(20), once we are given some \( \lambda \), it is observationally equivalent to estimate

\(^3\)Unless \( \tilde{\kappa}_i = 0 \) and, either \( \alpha_i = 0 \) or \( \beta_i = 0 \), for \( i = 1, 2, 3 \).
\( \theta_U = (s_0, (s_j)_{j=1,2,3}, \kappa_3, \mu_3, \alpha_3, \beta_3) \) or \( (\delta_0, (\delta_j)_{j=1,2,3}, \kappa_3, \mu_3, \alpha_3, \beta_3) \). We estimate \( \theta_U \), by creating moment conditions. We want to calibrate \( \theta_U \) in such a way that the model predicts patterns of ex-post returns and ex-post return volatility that match their empirical counterparts. We defer the empirical section for a careful description of the moment conditions we use to accomplish this task. For now, we define, for any given \( \lambda \),

\[
F_{T,J}(\theta_U(\lambda), \lambda) = \frac{1}{J} \sum_{j=1}^{J} f_j(\theta_U(\lambda), \lambda) - \frac{1}{T} \sum_{t=1}^{T} f_t,
\]

where \( f \) is a vector of moment conditions. Accordingly, we define, for any \( \lambda \), the probability limit \( \text{plim} \hat{\theta}_{U,T}(\lambda) \) [to be defined yet] = \( \theta_U^\dagger(\lambda) \), where

\[
\theta_U^\dagger(\lambda) = \arg \max_{\theta \in \Theta} E \left[ F_{T,J}(\theta_U(\lambda), \lambda) \cdot F_{T,J}(\theta_U(\lambda), \lambda) \right].
\]

We need the following assumptions:

**Assumptions:**

A1: \( E\nabla^2_{\theta} l_T(\theta_U^\dagger(\lambda), \lambda) \geq E\nabla^2_{\theta} l_T(\theta_U(\lambda), \lambda) \) for all \( \theta \neq \theta_U^\dagger \), for all \( \lambda \)

A2: \( E \left[ -\nabla^2_{\theta} l_T(\theta_U^\dagger(\lambda), \lambda) \right] \) uniformly positive definite in \( \theta \times \lambda \)

A3: \( E \left[ \left( \nabla_{\theta} l_T(\theta_U^\dagger(\lambda), \lambda) \right)^{4+\delta} \right] _i < \infty, i = 1, \ldots, k \), for some \( \delta > 0 \), where \( \theta \in \Theta \) with \( \Theta \) being a compact subset of \( \mathbb{R}^k \).

We have:

**Lemma 1** Let A(1)-A(3) hold. Then,

(i) \[
\sup_{\lambda \in \Lambda} \left| \hat{\theta}_{U,T}(\lambda) - \theta_U^\dagger(\lambda) \right| = o_P(1).
\]

(ii) Pointwise in \( \lambda \),

\[
\sqrt{T} \left( \hat{\theta}_{U,T}(\lambda) - \theta_U^\dagger(\lambda) \right) \stackrel{d}{\to} N \left( 0, B(\theta_O^\dagger(\lambda))^{-1} \right),
\]

where

\[
B(\theta_O^\dagger(\lambda))^{-1} = -E \left( \nabla_{\theta} \theta_T(\theta_O^\dagger, \lambda) \right) \quad \text{and} \quad A(\theta_O^\dagger(\lambda)) = \lim_{T \to \infty} \text{var} \left( \sqrt{T} \nabla_{\theta} \theta_T(\theta_O^\dagger, \lambda) \right).
\]

Note that we compute \( \hat{\theta}_T(\lambda) = \left( \hat{\theta}_{Q,T}, \hat{\theta}_{U,T}(\lambda) \right) \) through numerical approximations, as both the log-likelihood function and the objective functions in the SMM are not known in closed form. Accordingly, we do not dispose of a closed-form estimator of the asymptotic covariance matrix. Below, we shall see how to use bootstrap methods to overcome this issue.
3.2 Identifying the risk premium parameters

Given the risk premium parameters, the estimator for $\theta$ uses data on the two observable factors and the stock price to deliver estimates of all the parameters related to the dynamics of the three factors and the dividend process. We show how to use data on volatility contracts to identify and estimate the risk premium parameters, given the previous estimate of $\theta$. We aim to construct parametric and nonparametric measures of the conditional expectation of integrated volatility, under the risk neutral probability measure, i.e. $E[IV_{t,t+1}|F(t)]$. We identify $\lambda$ by minimizing a set of moment conditions based on the difference between the model-implied and model-free estimator of the risk neutral conditional expectation of integrated volatility.

Our first step involves deriving the model-implied estimator of $E[IV_{t,t+1}|F(t)]$. For this purpose, we need to simulate the price process under the risk-neutral measure. As the stock price is a deterministic affine function of the three factors, we need to simulate the factor dynamics under the risk neutral measure. Given our assumption that $\Sigma = I_{3 \times 3}$, this is given by

$$
dy(t) = [\kappa (\mu - y(t)) - V(y(t)) \Lambda(y(t))] dt + V(y(t)) d\hat{W}(t),
$$

which we simulate through a Milstein scheme, i.e. for each simulation $i = 1, \ldots, N$, and for $k = \xi^{-1}\tau$, we simulate $y(t)$ by drawing $y_{k:} = [y_{1,k:}, y_{2,k:}, y_{3,k:}]$, where $y_{k:} = y_{k:}(\theta)$, and for $\ell = 1, 2, 3$,

$$
y_{\ell,k:}^{(i)} = y_{\ell,k:}^{(i)} + \left[\kappa (\mu - y_{\ell,k:}^{(i)}) - V(y_{\ell,k:}^{(i)}) \Lambda(y_{\ell,k:}^{(i)})\right] \xi + \sqrt{\alpha_{\ell} + \beta_{\ell}y_{\ell,k:}^{(i)}} \epsilon_{\ell,k:}^{(i)}
- \frac{1}{2} \left[\frac{\partial}{\partial \theta} V(y_{\ell,k:}^{(i)}) \right]_{\xi}^2 + \frac{1}{2} \left[\frac{\partial}{\partial \theta} \sqrt{\alpha_{\ell} + \beta_{\ell}y_{\ell,k:}^{(i)}} \epsilon_{\ell,k:}^{(i)}\right]^2,
$$

where $\epsilon$ is NID with variance $\xi$, and assuming without loss of generality that $\xi^{-1}$ is an integer.

For each simulation $i$, we compute,

$$
\log s_{k:}^{(i)} = \log \left(s_0 + \sum_{\ell=1}^{3} s_{\ell,y_{\ell,k:}^{(i)}}\right).
$$

The processes in (23) and (24) are simulated for each value of $\lambda$, with parameter $\theta$ fixed at $\hat{\theta}_T(\lambda)$. We fix $J$ sets of initial values for the vector of factors, i.e. $y_j(0), j = 1, \ldots, J$. For each $j$, we simulate (22) and (24) using $\lambda$ and $\hat{\theta}_T(\lambda)$, and for each $\lambda$ we simulate $S$ paths of length one.\footnote{In practice we disregard a sufficiently high number of draws. However, for notational simplicity, we proceed as if we were to use $y_j(0)$ as the initial value.} For each initial value $y_j(0), \lambda \in \Lambda$, and for each simulation replication $i$, we thus obtain
log $s_{k\xi}^{(i)}(\hat{\theta}_T(\lambda),\lambda,y^j(0))$, for $k = 1, \ldots, \xi^{-1}$. For each $j = 1, \ldots, J$, define,

$$RV_{1,S} \left( \hat{\theta}_T(\lambda), \lambda, y^j (0) \right)$$

$$= \frac{1}{S} \sum_{i=1}^{S} \sum_{k=1}^{\xi-1} \left[ \log s_{(k+1)\xi}^{(i)} \left( \hat{\theta}_T(\lambda), \lambda, y^j (0) \right) - \log s_{k\xi}^{(i)} \left( \hat{\theta}_T(\lambda), \lambda, y^j (0) \right) \right]^2.$$  

For each $j$, as $S \to \infty$, $\xi \to 0$ and $S\xi \to \pi^{-1}$, for some positive constant $\pi < \infty$,

$$RV_{1,S} \left( \theta^i(\lambda), \lambda, y^j (0) \right) - \mathbb{E} \left[ IV_1 \left( \theta^i(\lambda), \lambda \right) \bigg| y(0) = y^j (0) \right] = O_P(S^{-1/2}),$$

and for $S/T \to 0$,

$$RV_{1,S} \left( \hat{\theta}_T(\lambda), \lambda, y^j (0) \right) - RV_{1,S} \left( \theta^i(\lambda), \lambda, y^j (0) \right) = O_P(T^{-1/2}).$$

Note that $RV_{1,S}((\hat{\theta}_T(\lambda), \lambda, y^j(0)))$ is an estimator of $\mathbb{E} \left[ IV_1 \left( \theta^i(\lambda), \lambda \right) \bigg| y(0) = y^j (0) \right]$, i.e. it is a model-implied estimator of the expected integrated volatility, under the risk-neutral measure, conditional on $y(0) = y^j(0)$.

Next, we turn to the model-free estimator of $\mathbb{E}[IV_{t+1}F(t)]$ introduced in Section 2.4 (see Eq. (13)). In practice, we only observe a discrete number of strike prices. Therefore, we have to go through a numerical approximation to Eq. (13),

$$\mathbb{E}[IV_{t+1}F(t)] = \frac{2}{N} \left[ \sum_{i:K_i \leq F(t)} \frac{P(t,t+1,K_i)}{u(t,t+1)} K_i + \sum_{i:K_i > F(t)} \frac{C(t,t+1,K_i)}{u(t,t+1)} K_i \right],$$

for $t = 1, \cdots, T$.\footnote{Jiang and Tian (2005) have demonstrated that the approximation of a Riemann integral of OTM put and call option prices with a Riemann sum works very well in practice.} Finally, define

$$G_{T,S,J,N}(\hat{\theta}_T(\lambda), \lambda) = \frac{1}{J} \sum_{j=1}^{J} g_j.S(\hat{\theta}_T(\lambda), \lambda) - \frac{1}{T} \sum_{t=1}^{T} g_{t,N}$$

where $(1/J) \sum_{j=1}^{J} g_j.S(\hat{\theta}_T(\lambda), \lambda)$ is a vector containing $q$, $q \geq 3$, moment conditions constructed using $RV_{1,S}((\hat{\theta}_T(\lambda), \lambda, y^j(0)))$, and $(1/T) \sum_{t=1}^{T} g_{t,N}$ is a vector containing $q$ moment conditions constructed using $\mathbb{E}[IV_{t+1}F(t)]$ in Eq. (25).

Note that, as we are averaging over the conditioning variables, we are indeed matching unconditional moments of model-implied integrated volatility to their sample counterparts, obtained with model-free volatility, where moments are computed under the risk neutral probability measure. For example, we can compare mean, variance and first and $k$th autocovariance of model implied and model free integrated volatility, under the risk neutral probability measure. As shown below, $\lambda$ can be estimated by minimizing the distance between moment conditions based on parametric and nonparametric estimator.
3.3 A first step estimator of $\lambda$

We estimate $\lambda$ using Simulated-GMM (SGMM). Define,

$$\hat{\lambda}_{T,S,J,N} = \arg \min_{\lambda} G_{T,S,J,N}(\hat{\theta}_T(\lambda), \lambda)^T W_{T,N}^{-1} G_{T,S,J,N}(\hat{\theta}_T(\lambda), \lambda) \equiv \arg \min_{\lambda} Z_{T,S,J,N}(\lambda),$$

where $W_{T,N}^{-1}$ is a weighting function constructed using $\mathbb{E}[IV_{t,t+1}|\mathcal{F}(t)]$. Define, also,

$$Z_{T,S,J,N}^1(\lambda) = G_{T,S,J,N}(\theta_T^*(\lambda), \lambda)^T W_{T,N}^{-1} G_{T,S,J,N}(\theta_T^*(\lambda), \lambda)$$

and

$$\lambda^* = \arg \min_{\lambda} Z_{\infty}(\lambda), \text{ where } Z_{\infty}(\lambda) = \text{plim}_{T,S,J,N} Z_{T,S,J,N}^1(\lambda).$$

The difficulty lies in the fact that the moment conditions depends on estimated parameters, i.e. on $\hat{\theta}_T(\lambda)$, where the latter is a square-root consistent estimator. Therefore, the limiting distribution will reflect the contribution of parameter estimation error. In particular, the asymptotic covariance of $\sqrt{T}G_{T,S,J,N}(\hat{\theta}_T(\lambda^*), \lambda^*)$ will reflect the contribution of the asymptotic covariance of $\sqrt{T}(\hat{\theta}_T(\lambda) - \theta_T^*(\lambda))$, which is not known in closed form.

Consider the following conditions.

**Assumptions:**

A4: $Z_{\infty}(\lambda) \leq Z_{\infty}(\lambda')$ for all $\lambda \neq \lambda'$.

A5: $W_{T,N}^{-1} - W_{\infty}^{-1} = o_P(1)$.

We have:

**Lemma 2** Let A1-A5 hold. If $\xi^{-1}/S \to \pi$, $\pi > 0$, $N/T \to \infty$, $S/T \to \infty$, and $J/T \to \infty$, then,

$$\sqrt{T}G_{T,S,J,N}(\hat{\theta}_T(\lambda^*), \lambda^*) \overset{d}{=} N(0, V),$$

where

$$V = \lim_{T \to \infty} \text{var} \left[ A_{\theta}^T \sqrt{T} (\hat{\theta}_T(\lambda^*) - \theta_T^*(\lambda^*)) \right] + \lim_{T,S,J,N \to \infty} \text{var} \left[ \sqrt{T}G_{T,S,J,N}(\theta_T^*(\lambda^*), \lambda^*) \right]$$

$$+ \lim_{T,S,J,N \to \infty} \text{cov} \left[ \sqrt{T}G_{T,S,J,N}(\theta_T^*(\lambda^*), \lambda^*), A_{\theta}^T \sqrt{T} (\hat{\theta}_T(\lambda) - \theta_T^*(\lambda^*)) \right], \quad (26)$$

and $A_{\theta}^T = \text{plim}_{T,S,J,N} \nabla_{\theta} G_{T,S,J,N}(\theta_T^*(\lambda^*), \lambda^*)$.

It is immediate to see that the limiting distribution of $\sqrt{T}G_{T,S,J,N}(\hat{\theta}_T(\lambda^*), \lambda^*)$ reflects the contribution of the asymptotic covariance of $\sqrt{T}(\hat{\theta}_T(\lambda) - \theta_T^*(\lambda))$, which is not known in closed form. For this reason, $W_{T,N}$ cannot be an optimal weighting matrix. In fact, $W_{\infty} = \text{plim}_{T,N} W_{T,N}$ is the inverse of the long-run variance of $\sqrt{T}G_{T,S,J,N}(\theta_T^*(\lambda^*), \lambda^*)$ which differs from the long-run variance of $\sqrt{T}G_{T,S,J,N}(\hat{\theta}_T(\lambda^*), \lambda^*)$. As a consequence, $\hat{\lambda}_{T,S,J,N}$ cannot be an efficient estimator. In fact,
Theorem 3 Let A1-A5 hold. If $\xi^{-1}/S \to \pi$, $\pi > 0$, $N/T \to \infty$, $S/T \to \infty$, and $J/T \to \infty$, then,

$$
\sqrt{T}(\hat{\theta}_{T,S,J,N} - \lambda^1) \overset{d}{\to} N \left(0, (A^\dagger_{\lambda^1} W_{\infty}^{-1} A^\dagger_{\lambda^1})^{-1} A^\dagger_{\lambda^1} W_{\infty}^{-1} V W_{\infty}^{-1} A^\dagger_{\lambda^1} \left(A^\dagger_{\lambda^1} W_{\infty}^{-1} A^\dagger_{\lambda^1}\right)^{-1}\right),
$$

where $A^\dagger_{\lambda^1} = \text{plim}_{T,S,J,N \to \infty} \nabla G_{T,S,J,N}(\theta^\dagger(\lambda^1), \lambda^1)$, and $V$ is as in Lemma 2.

3.4 Bootstrap Optimal Weighting Matrix

To obtain an efficient estimator, we need to use a weighting matrix which converges in probability to $V^{-1}$. The critical issue, however, is that there is no available estimator of $V$, as

$$
\var\left[\left(A^\dagger_{\theta^1} \sqrt{T} \left(\hat{\theta}_T(\lambda^1) - \theta^\dagger(\lambda^1)\right)\right)\right]
$$

is not known in closed-form. We outline a bootstrap procedure for constructing an optimal weighting matrix.

Within the class of affine models, $IV_t$ is a strong mixing process, and so $E[IV_{t,t+1}|F_t]$ is also strong mixing, provided $F_t$ represents a finite history of information. To capture the correlation structure in the conditional expectation of integrated volatility, we need to rely on a block bootstrap procedure. Furthermore, in order to capture the contribution of the covariance term in (26), we need to jointly resample blocks of the likelihood function and of the conditional expectation of integrated volatility.

Resample $b$ blocks of length $l$, $lb = T$, of $w_t(\theta, \lambda) = (l_t(\theta, \lambda), g_t(\theta, \lambda))$ so that

$$
w_1^\ast(\theta, \lambda), \ldots, w_l^\ast(\theta, \lambda), w_{l+1}^\ast(\theta, \lambda), \ldots, w_T^\ast(\theta, \lambda)
$$

is equal to

$$
w_{I_1}(\theta, \lambda), \ldots, w_{I_1+i}(\theta, \lambda), w_{I_2}(\theta, \lambda), \ldots, w_{I_b+l}(\theta, \lambda),
$$

where $I_i, i = 1, \ldots, b$ are discrete i.i.d. uniform on $0, 1, \ldots, T - l - 1$. Define,

$$
\hat{\theta}^\ast_T(\hat{\lambda}_{T,S,J,N}) = \arg\max_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^T l_t^\ast(\theta, \hat{\lambda}_{T,S,J,N}),
$$

and

$$
G^\ast_{T,S,J,N}(\hat{\theta}^\ast_T(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N}) = \frac{1}{T} \sum_{t=1}^T \left(\frac{1}{J} \sum_{j=1}^J g_{j,S}(\hat{\theta}^\ast_T(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N}) - g_{i,N}^\ast\right).
$$

$^6$Note that if the affine diffusion is indeed correctly specified for the factor dynamics and if the no arbitrage condition hold, then $l_t(\theta, \lambda)$ is i.i.d. for all $\theta$ and $\lambda$.  

16
Note that we have not resampled \( g_{j,S}(\theta(\lambda), \lambda) \); in fact for \( S/T \to \infty \) and \( J/T \to \infty \), the simulation error is negligible, and
\[
\sqrt{T} \left( \frac{1}{J} \sum_{j=1}^{J} g_{j,S}(\theta(\lambda), \lambda) - E(g_{j,S}(\theta(\lambda), \lambda)) \right) = o_p(1).
\]
Nevertheless, in order to properly capture the contribution of parameter estimation error, the model based moment conditions are evaluated at the bootstrap parameters. Also, define
\[
G_{T,S,J,N}(\hat{\theta}_{T}(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N}) = \frac{1}{T} \sum_{t=1}^{T} \left( \frac{1}{J} \sum_{j=1}^{J} g_{j,S}(\hat{\theta}_{T}(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N}) - g_{t,N} \right),
\]
and note that it denotes the mean of the bootstrap moment conditions, which is of order \( O_p(T^{-1/2}) \) in the overidentified case.\(^7\) Therefore, in the overidentified case, recentering is necessary even for first order validity.

Basically, it can be shown that the limiting distribution of
\[
\sqrt{T} \left[ G_{T,S,J,N}^*(\hat{\theta}_{T}^*(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N}) - G_{T,S,J,N}(\hat{\theta}_{T}(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N}) \right]
\]
is the same as the limiting distribution of \( \sqrt{T} G_{T,S,J,N}(\hat{\theta}_{T}(\lambda^1), \lambda^1) \), conditional on the sample and for all sample but a set of measure approaching zero. It is well known that convergence in distribution does not necessarily imply convergence of moments, however if the \((2 + \delta)\)-th moments of the bootstrap statistic is finite, then
\[
\text{var}^* \left[ \sqrt{T} G_{T,S,J,N}^*(\hat{\theta}_{T}^*(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N}) \right],
\]
as \( T, S, J, N \to \infty \), will approach
\[
\lim_{T,S,J,N \to \infty} \text{var} \left[ \sqrt{T} G_{T,S,J,N}(\hat{\theta}_{T}(\lambda^1), \lambda^1) \right].
\]
Hence, it is sufficient to construct \( B \) bootstrap statistics, with \( B \) large enough, and compute their sample variance (see e.g. Goncalves and White, 2005, for the least squares case). Thus,
\[
V_{T,B}(\hat{\theta}_{T}(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N})
\]
\[
= \frac{T}{B} \sum_{i=1}^{B} \left( G_{T,S,J,N,i}^*(\hat{\theta}_{T,i}^*(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N}) - G_{T,S,J,N}(\hat{\theta}_{T}(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N}) \right)^\top 
\]
\[
\left( G_{T,S,J,N,i}^*(\hat{\theta}_{T,i}^*(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N}) - G_{T,S,J,N}(\hat{\theta}_{T}(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N}) \right),
\]
\(^7\)In the exactly identified case, \( G_{T,S,J,N}(\hat{\theta}_{T}(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N}) \) is instead \( O_p(l/T) \), and so recentering is not necessary for the first order validity of the bootstrap.
and as $B \to \infty$, $V_{T,B} \left( \hat{\theta}_T(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N} \right)$ approaches

$$\text{var}^* \left[ \sqrt{T} G_{T,S,J,N}^* \left( \hat{\theta}_T^* (\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N} \right) \right].$$

**Theorem 4** Let $A1$-$A5$ hold. If $S/J \to \infty$, $\xi^{-1} / S \to \infty$, $J/T \to \infty$, $N/T \to \infty$ and $l/T^{1/2} \to 0$, then as $T \to \infty$ and $B \to \infty$,

$$P \left( \omega : P^* \left( \left| V_{T,B} \left( \hat{\theta}_T(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N} \right) - V \right| > \varepsilon \right) \right) \to 0,$$

where $P^*$ is the probability law governing the bootstrap and $V$ is as in Lemma 2.

Therefore, the optimal weighting matrix is given by $V_{T,B} \left( \hat{\theta}_T(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N} \right)^{-1}$.

### 3.5 An efficient Concentrated SGMM estimator

By using $V_{T,B} \left( \hat{\theta}_T(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N} \right)^{-1}$ as weighting matrix we can obtain an efficient estimator. Define,

$$\tilde{\lambda}_{T,S,J,N} = \arg \min_{\lambda \in \Lambda} G_{T,S,J,N} \left( \hat{\theta}_T (\lambda), \lambda \right)^\top V_{T,B} \left( \hat{\theta}_T(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N} \right)^{-1} G_{T,S,J,N} \left( \hat{\theta}_T (\lambda), \lambda \right).$$

We have:

**Lemma 5** Let $A1$-$A5$ hold. If $\xi^{-1} / S \to \pi$, $\pi > 0$, $N/T \to \infty$, $S/T \to \infty$, $l/T^{1/2} \to 0$ and $J/T \to \infty$, then as $B \to \infty$,

$$\sqrt{T} \left( \tilde{\lambda}_{T,S,J,N} - \lambda^1 \right) \overset{d}{\rightarrow} N \left( 0, \left( A_{\lambda^1}^{\top} V A_{\lambda^1} \right)^{-1} \right),$$

where $A_{\lambda^1}^{\top} = \text{plim}_{T,S,J,N \to \infty} \nabla_{\lambda} G_{T,S,J,N} \left( \theta^1 (\lambda^1), \lambda^1 \right)$, and $V$ is as in Lemma 2.

Though from Theorem 3 we know how to consistently estimate $V$, still we do not have a closed form expression for $A_{\lambda^1}^{\top}$, and so a ready-to-use estimator is not available. We can overcome this problem computing bootstrap critical values for $\sqrt{T} \left( \tilde{\lambda}_{T,S,J,N} - \lambda^1 \right)$. Define,

$$\tilde{\lambda}_{T,S,J,N}^* = \arg \min_{\lambda \in \Lambda} \left[ \left( G_{T,S,J,N}^* \left( \hat{\theta}_T^* (\lambda), \lambda \right) - G_{T,S,J,N} \left( \hat{\theta}_T(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N} \right) \right) \right]^\top V_{T,B} \left( \hat{\theta}_T(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N} \right)^{-1} \left( G_{T,S,J,N}^* \left( \hat{\theta}_T^* (\lambda), \lambda \right) - G_{T,S,J,N} \left( \hat{\theta}_T(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N} \right) \right).$$

We have,
Theorem 6  Let A1-A5 hold. If \( S/J \rightarrow \infty, \xi^{-1}/S \rightarrow \infty, J/T \rightarrow \infty, N/T \rightarrow \infty, l/T^{1/2} \rightarrow 0 \) and \( l/T^{1/2} \rightarrow 0 \), then as \( T \rightarrow \infty \) and \( B \rightarrow \infty \),

\[
P\left( \omega : \sup_{x \in \mathbb{R}} P^* \left[ \sqrt{T} \left( \tilde{\lambda}_{T,S,J,N}^* - \tilde{\lambda}_{T,S,J,N} \right) \leq x \right] - P \left[ \sqrt{T} \left( (\tilde{\lambda}_{T,S,J,N} - \lambda^t) \right) \leq x \right] > \varepsilon \right) \rightarrow 0,
\]

where \( P^* \) is the probability law governing the bootstrap.

4  Empirical analysis

4.1  Data

Our sample data include the consumer price index and the index of industrial production for the US, observed monthly from January 1950 to December 2006, for a total of 672 observations. We take these two series to compute the two macroeconomic factors, the gross inflation and the gross industrial production growth, both at a yearly level,

\[
y_{1,t} \equiv \frac{\text{CPI}_t}{\text{CPI}_{t-12}} \quad \text{and} \quad y_{2,t} \equiv \frac{\text{IP}_t}{\text{IP}_{t-12}},
\]

where CPI\(_t\) is the consumer price index and IP\(_t\) is the seasonally adjusted industrial production index, as of month \( t \). Figure 1 depicts the two series \( y_{1,t} \) and \( y_{2,t} \), along with NBER-dated recession events. As for the stock price data, we use the S&P Compounded index and the VIX index. Data for the VIX index are available daily, but only for the period following January 1990. Information related to the CPI and the IP is made available to the market between the 19th and the 23rd of every month. Thus, to possibly avoid overreaction to releases of information, we sample the S&P Compounded index and the VIX index every 25th of the month.

4.2  Moment conditions for the stock price process

To form moment conditions related to the stock price process, we match key moment conditions related to the dynamics of returns and return volatility. First, we compute the continuously compounded price changes as, \( R_t = \log (s_t/s_{t-12}) \), where \( s_t \) denotes the real stock price, obtained as the ratio between the S&P and the CPI. Following Mele (2007) and Fornari and Mele (2006), we measure the volatility of the continuously compounded price changes, \( R_t \), as,

\[
\text{Vol}_t = \sqrt{6\pi} \cdot \frac{1}{12} \sum_{i=1}^{12} \left| R_{t+1-i}^{m} \right|,
\]

where \( R_{t}^{m} \) is the monthly, continuously compounded price change, defined as \( R_{t}^{m} = \log (s_t/s_{t-1}) \).

Second, we consider the following regressions,

\[
R_t = a^R + b_{1,12}^R y_{1,t-12} + b_{2,12}^R y_{2,t-12} + \epsilon_t^R,
\]

19
and,

\[ \text{Vol}_t = a^V + \sum_{i \in \{6,12,18,24,36,48\}} \phi_i \text{Vol}_{t-i} + \sum_{i \in \{12,24,36,48\}} b_{1,i}^V y_{1,t-i} + \sum_{i \in \{12,24,36,48\}} b_{2,i}^V y_{2,t-i} + \epsilon_t^V, \]

where \( \epsilon_t^R \) and \( \epsilon_t^V \) are residual terms. Third, define \( \mu_S \) as the average of the real stock price, \( s \), and \( V_S \) to be the average of return volatility, \( \text{Vol} \).

The parameter vector we match is,

\[ \psi_{R,V} = \left( a^R, b_{1,12}^R, b_{2,12}^R, a^V, (\phi_i)_{i \in \{6,12,18,24,36,48\}}, (b_{1,i}^V)_{i \in \{12,24,36,48\}}, (b_{2,i}^V)_{i \in \{12,24,36,48\}}, \mu_S, V_S \), \]

and we fix the parameter \( s_0 \) to,

\[ s_0 = \mu_S - s_1 \bar{y}_1 - s_2 \bar{y}_2 - s_3 \mu_3, \]

where \( \bar{y}_1 \) and \( \bar{y}_2 \) are the sample average of \( y_{1,t} \) and \( y_{2,t} \).

### 4.3 Moment conditions for the risk-adjusted stock volatility

We consider the following regression,

\[ \text{VIX}_t = a^{\text{VIX}} + \sum_{i \in \{1,12,24,36\}} b_{1,i}^{\text{VIX}} y_{1,t-i} + \sum_{i \in \{1,12,24,36\}} b_{2,i}^{\text{VIX}} y_{2,t-i} + \epsilon_t^{\text{VIX}}, \]

where \( \text{VIX}_t \) is the VIX index as of month \( t \), and \( \epsilon_t^{\text{VIX}} \) is a residual term. We match the following parameter vector,

\[ \psi^{\text{VIX}} = \left( a^{\text{VIX}}, (b_{1,i}^{\text{VIX}})_{i \in \{1,12,24,36\}}, (b_{2,i}^{\text{VIX}})_{i \in \{1,12,24,36\}}, \mu^{\text{VIX}}, \sigma^{\text{VIX}} \right), \]

where \( \mu^{\text{VIX}} \) and \( \sigma^{\text{VIX}} \) are the average and the standard deviation of the VIX index.

### 4.4 Estimation results

Table 1 reports moment conditions.

Tables 2 through 4 report parameter estimates.

Figure 2 depicts sample data related to the continuously compounded price changes, \( R_t \), return volatility, \( \text{Vol}_t \), along with the dynamics predicted by the model. The third panel in Figure 2 plots the instantaneous volatility implied by the model.

Figure 3 depicts the VIX index, along with the VIX predicted by the model, the (square root of the) model-implied expected integrated variance, and the instantaneous volatility predicted by the model. The second panel in Figure 2 plots the volatility risk-premium, defined as the
difference between the (square roots of the) model-implied expected integrated variance under the risk-neutral probability and the model-implied expected integrated variance under the physical probability.

Figure 4 provides scatterplots of the volatility risk-premium against inflation and industrial production.

5 Conclusion

This paper develops a framework to analyze the business cycle movements of stock market returns, volatility and volatility risk-premia. In our model, the aggregate stock market behavior relates to the development of two macroeconomic factors, inflation and industrial production growth, and one unobserved factor. The relations linking the asset price, returns and volatility to these factors are derived under the assumption of no-arbitrage. This key aspect differentiates our approach from previous models with stochastic volatility, in which volatility was specified exogenously to the price process.

We take our model to data, and make use of the new volatility index, VIX, to estimate the parameters related to risk-aversion. Our model predicts that stock market returns are procyclical, stock market volatility is countercyclical and volatility risk-premia are countercyclical.
Appendix

A. Proofs for Section 2

Existence of a strong solution to (1). Consider the following conditions: For all $i$,

(i) For all $y_i$, $V(y_i) = 0$, $\beta_i > \Sigma > \beta_i$

(ii) For all $j$, if $(\beta_j \Sigma)_{ij} = 0$, then $V_{ii} = V_{jj}$.

Then, by Duffie and Kan (1996) (unnumbered theorem, p. 388), there exists a unique strong solution to Eq. (1) for which $V(y(t)) > 0$ for all $t$ almost surely.

We apply these conditions to the case in which $\Sigma = I_{3 \times 3}$, $\beta_i$ is a vector of zeros, except possibly for its $i$-th element, denoted as $\beta_i \equiv \beta_{ii}$, and $\kappa$ is as in Section 2.5. Condition (i) collapses to,

For all $y_i : \alpha_i + \beta_i y_i = 0$,  $\beta_i \left( \frac{\alpha_i}{\beta_j} \right) > \frac{1}{2} \beta_i^2$, $i \neq j, i, j \in \{1, 2\}$

That is, ruling out the trivial case $\beta_i = 0$.

For all $y_3 : \alpha_3 + \beta_3 y_3 = 0$, $\beta_3 \alpha_3 = \frac{1}{2} \beta_3^2$

Proof of Proposition 1. Define the Arrow-Debreu adjusted asset price process as, $s^\xi(t) = e^{-rt} \xi(t) s(y(t))$, $t > 0$. By Itô’s lemma, it satisfies,

$$\frac{ds^\xi(t)}{s^\xi(t)} = D(y(t)) dt + \left( Q(y(t)) - \Lambda(y(t)) \right) dW(t), \quad (A1)$$

where

$$D(y) = -r + \frac{As(y)}{s(y)} - Q(y) \Lambda(y); \quad Q(y) = \frac{s_y(y)^T \Sigma V(y)}{s(y)}$$

$$As(y) = s_y(y)^T \kappa (\mu - y) + \frac{1}{2} \text{Tr} \left( \Sigma V(y) [\Sigma V(y)]^T s_yy(y) \right).$$

By absence of arbitrage opportunities, for any $T < \infty$,

$$s^\xi(T) = E \left[ \int_t^T \delta^\xi(h) \, dh \, | \, F(t) \right] + E[s^\xi(T) | F(t)], \quad (A2)$$
where $\delta^\xi(t)$ is the current Arrow-Debreu value of the dividend to be paid off at time $t$, viz $\delta^\xi(t) = e^{-rt}\xi(t)\delta(t)$. Below, we show that the following transversality condition holds,

$$\lim_{T \to \infty} E[s^\xi(T) \mid F(t)] = 0,$$

from which Eq. (7) in the main text follows.

Next, by Eq. (A2),

$$0 = \frac{d}{d\tau} E[s^\xi(\tau) \mid F(t)] \bigg|_{\tau=t} + \delta^\xi(t).$$

Below, we show that

$$E[s^\xi(T) \mid F(t)] = s^\xi(t) + \int_t^T D(y(h)) s^\xi(h) dh.$$

Therefore, by the assumptions on $\Lambda$, Eq. (A4) can be rearranged to yield the following partial differential equation,

For all $y$, $s_y(y)^\top(c - Dy) + \frac{1}{2} \text{Tr} \left( [\Sigma V(y)] [\Sigma V(y)]^\top s_{yy}(y) \right) + \delta(y) - rs(y) = 0,$

(A6)

where $c$ and $D$ are defined in the proposition.

Let us assume that the price function is affine in $y$,

$$s(y) = \gamma + \eta^\top y,$$

(A7)

for some scalar $\gamma$ and some vector $\eta$. By plugging this guess back into Eq. (A6) we obtain,

For all $y$, $\eta^\top c + \delta_0 - r\gamma - \left[ \eta^\top(D + rI_{n \times n}) - \delta^\top \right] y = 0.$

That is,

$$\eta^\top c + \delta_0 - r\gamma = 0 \quad \text{and} \quad \left[ \eta^\top(D + rI_{n \times n}) - \delta^\top \right] = 0_{1 \times n}.$$

The solution to this system is,

$$\gamma = \frac{\delta_0 + \eta^\top c}{r} \quad \text{and} \quad \eta^\top = \delta^\top (D + rI_{n \times n})^{-1}.$$

We are left to show that Eq. (A3) and (A5) hold true.

As regards Eq. (A3), we have

$$\lim_{T \to \infty} E[s^\xi(T) \mid F(t)] = \lim_{T \to \infty} E[e^{-r(T-t)}\xi(T) s(y(T)) \mid F(t)]$$

$$= \gamma e^{-r(T-t)} \lim_{T \to \infty} E[\xi(T) \mid F(t)] + \lim_{T \to \infty} e^{-r(T-t)} E[\xi(T) \eta^\top y(T) \mid F(t)]$$

$$= \xi(t) \lim_{T \to \infty} e^{-r(T-t)} E[\eta^\top y(T) \mid F(t)],$$

23
where the second line follows by Eq. (A7), and the third line holds because \( E[\xi(T) \mid \mathbb{F}(t)] = 1 \) for all \( T \), and by a change of measure (Warning: We need some more work to prove this.) Eq. (A3) follows because \( y \) is stationary mean-reverting.

To show that Eq. (A5) holds, we need to show that the diffusion part of \( s^\xi \) in Eq. (A1) is a martingale, not only a local martingale, which it does whenever for all \( T \),

\[
\int_T^\infty E[\xi(T) \mid \mathbb{F}(t)] \left( \xi(T) - \xi(t) \right) dt = 1
\]

for all \( T \), and by a change of measure (Warning: We need some more work to prove this.) Eq. (A3) follows because \( y \) is stationary mean-reverting.

\[\Box\]

**B. Proofs for Section 3**

**Proof of Lemma A1:** (i) By the uniform law of large numbers,

\[
\sup_{\theta \in \Theta} \sup_{\lambda \in \Lambda} |l_T(\theta, \lambda) - E(l_T(\theta, \lambda))| = o_P(1).
\]

Given A1, the desired outcome follows from Lemma A1 in Andrews (1993).

(ii) Via a mean value expansion around \( \theta^* (\lambda) \),

\[
0 = \nabla_\theta l_T(\hat{\theta}_T(\lambda)) = \nabla_\theta l_T(\theta^\dagger(\lambda)) + \nabla_\theta^2 l_T(\bar{\theta}_T(\lambda)) \left( \bar{\theta}_T(\lambda) - \theta^\dagger(\lambda) \right)
\]

with \( \bar{\theta}_T(\lambda) \in (\hat{\theta}_T(\lambda), \theta^*(\lambda)) \), and so

\[
\sqrt{T} \left( \bar{\theta}_T(\lambda) - \theta^\dagger(\lambda) \right) = B(\theta^\dagger(\lambda))^{-1} \nabla_\theta l_T(\theta^\dagger(\lambda)) + o_P(1)
\]

where the \( o_P(1) \) holds uniformly in \( \lambda \), given that \( \sup_{\theta \times \lambda \in \Theta \times \Lambda} |\nabla_\theta^2 l_T(\theta(\lambda)) - B(\theta^*(\lambda))| = o_P(1) \). Given that the affine model considered is stationary ergodic, and so geometrically strong mixing, the statement follows by the central limit theorem.

**Proof of Lemma A2:**

\[
\sqrt{T} G_{T,S,J,N}(\hat{\theta}_T(\lambda^\dagger), \lambda^\dagger) = \sqrt{T} G_{T,S,J,N}(\theta^\dagger(\lambda^\dagger), \lambda^\dagger)
\]

\[
+ \nabla_\theta G_{T,S,J,N}(\bar{\theta}_T(\lambda^\dagger), \lambda^\dagger) \sqrt{T} \left( \bar{\theta}_T(\lambda^\dagger) - \theta^\dagger(\lambda^\dagger) \right)
\]

\[
= \sqrt{T} G_{T,S,J,N}(\theta^\dagger(\lambda^\dagger), \lambda^\dagger) + A^\dagger_{\theta} \sqrt{T} \left( \hat{\theta}_T(\lambda^\dagger) - \theta^\dagger(\lambda^\dagger) \right) + o_P(1),
\]

given that, by Lemma A1,

\[
\nabla_\theta G_{T,S,J,N}(\bar{\theta}_T(\lambda^\dagger), \lambda^\dagger) = \nabla_\theta G_{T,S,J,N}(\theta^\dagger(\lambda^\dagger), \lambda^\dagger) + o_P(1)
\]

24
and \( \text{plim}_{T,S,J,N \rightarrow \infty} \nabla_\theta G_{T,S,J} (\theta^\dagger (\lambda^l), \lambda^l) = A_{\theta^l}^\dagger \). The result then follows from central limit theorem.

**Proof of Theorem 3:** Via a mean value expansion around \( \lambda^* \),

\[
0 = \sqrt{T} \nabla_\lambda \tilde{Z}_{T,S,J,N} (\tilde{\lambda}_{T,S,J,N})
= \nabla_\lambda G_{T,S,J,N} (\hat{\theta}_T (\tilde{\lambda}_{T,S,J,N}), \tilde{\lambda}_{T,S,J,N}) \nabla_\lambda G_{T,S,J,N} (\hat{\theta}_T (\tilde{\lambda}_{T,S,J,N}), \tilde{\lambda}_{T,S,J,N})^{-1} \nabla_\lambda G_{T,S,J,N} (\hat{\theta}_T (\tilde{\lambda}_{T,S,J,N}), \tilde{\lambda}_{T,S,J,N})
\]
\[\times \sqrt{T} \left( \tilde{\lambda}_{T,S,J,N} - \lambda^l \right) .
\]

Thus,

\[
\sqrt{T} \left( \tilde{\lambda}_{T,S,J,N} - \lambda^l \right)
= - \left( \nabla_\lambda G_{T,S,J,N} (\hat{\theta}_T (\tilde{\lambda}_{T,S,J,N}), \tilde{\lambda}_{T,S,J,N}) \nabla_\lambda G_{T,S,J,N} (\hat{\theta}_T (\tilde{\lambda}_{T,S,J,N}), \tilde{\lambda}_{T,S,J,N}) \right)^{-1}
\times \nabla_\lambda G_{T,S,J,N} (\hat{\theta}_T (\tilde{\lambda}_{T,S,J,N}), \tilde{\lambda}_{T,S,J,N}) \nabla_\lambda G_{T,S,J,N} (\hat{\theta}_T (\tilde{\lambda}_{T,S,J,N}), \tilde{\lambda}_{T,S,J,N})
\]
\[\times \sqrt{T} (\tilde{\lambda}_{T,S,J,N} - \lambda^l) .
\]

Given Theorem 1 and Lemma A1,

\[
\text{plim}_{T,S,J,N \rightarrow \infty} \nabla_\lambda G_{T,S,J} (\hat{\theta}_T (\tilde{\lambda}_{T,S,J}), \tilde{\lambda}_{T,S,J,N}) \nabla_\lambda G_{T,S,J} (\hat{\theta}_T (\tilde{\lambda}_{T,S,J}), \tilde{\lambda}_{T,S,J,N})^{-1} \nabla_\lambda G_{T,S,J} (\hat{\theta}_T (\tilde{\lambda}_{T,S,J}), \tilde{\lambda}_{T,S,J,N})
\]
\[\times \sqrt{T} (\tilde{\lambda}_{T,S,J,N} - \lambda^l) .
\]

The statement then follows from Lemma A2.

**Proof of Theorem 4:** The proof of the Lemma is based on two main steps.

**Step 1:** We show that,

\[
P \left( \omega : \sup_{x \in \mathbb{R}} \left[ P^* \left[ \sqrt{T} \left( G_{T,S,J,N}^* (\hat{\theta}_T (\tilde{\lambda}_{T,S,J,N}), \tilde{\lambda}_{T,S,J,N}) - G_{T,S,J,N} (\hat{\theta}_T (\tilde{\lambda}_{T,S,J,N}), \tilde{\lambda}_{T,S,J,N}) \right) \right] \leq x \right] \right)
\]
\[\sup_{u,v \in U \times V} \sup_{u,v \in U \times V} \left[ P^* \left[ \sqrt{T} (G_{T,S,J,N}^* (\hat{\theta}_T (\tilde{\lambda}_{T,S,J,N}), \tilde{\lambda}_{T,S,J,N}) - G_{T,S,J,N} (\hat{\theta}_T (\tilde{\lambda}_{T,S,J,N}), \tilde{\lambda}_{T,S,J,N})) \leq x \right] \right] \rightarrow 0
\]

**Step 2:** We show that,

\[
E^* \left[ \left( \sqrt{T} (G_{T,S,J,N}^* (\hat{\theta}_T (\tilde{\lambda}_{T,S,J,N}), \tilde{\lambda}_{T,S,J,N}) - G_{T,S,J,N} (\hat{\theta}_T (\tilde{\lambda}_{T,S,J,N}), \tilde{\lambda}_{T,S,J,N})) \right)^{2+\delta} \right] \leq 0
\]

where \( E^* \) denotes the expectation according to the bootstrap probability law, conditional on the sample.

Given the statement in Steps 1 and 2, by the Corollary to Theorem 25.12 in Billingsley (1986), it follows that as \( T \rightarrow \infty \),

\[
\text{var}^* \left[ \sqrt{T} (G_{T,S,J,N}^* (\hat{\theta}_T (\tilde{\lambda}_{T,S,J,N}), \tilde{\lambda}_{T,S,J,N}) - G_{T,S,J,N} (\hat{\theta}_T (\tilde{\lambda}_{T,S,J,N}), \tilde{\lambda}_{T,S,J,N})) \right] \leq \mathbf{V},
\]

25
where \( \text{var}^* \) denotes the variance according to the bootstrap probability law, conditional on the sample. As \( B \to \infty \),

\[
V_{T,B} \left( \hat{\theta}_T(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N} \right) \xrightarrow{p} \text{var}^* \left[ \sqrt{T} \left( G^*_{T,S,J,N} \left( \hat{\theta}_T^*(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N} \right) - G_{T,S,J,N} \left( \hat{\theta}_T(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N} \right) \right) \right],
\]

the statement above follows.

**Proof of Step 1:**

\[
G^*_{T,S,J,N} \left( \hat{\theta}_T^*(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N} \right) - G_{T,S,J,N} \left( \hat{\theta}_T(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N} \right)
\]

\[
= G^*_{T,S,J,N} \left( \hat{\theta}_T(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N} \right) - G_{T,S,J,N} \left( \hat{\theta}_T(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N} \right)
\]

\[
+ \nabla_{\theta} G^*_{T,S,J,N} \left( \hat{\theta}_T^*(\hat{\lambda}_{T,S,J,N}), \hat{\lambda}_{T,S,J,N} \right) \sqrt{T} \left[ \hat{\theta}_T^*(\hat{\lambda}_{T,S,J,N}) - \hat{\theta}_T(\hat{\lambda}_{T,S,J,N}) \right]
\]

We begin by showing that

\[
P \left( \omega : \sup_{x \in \mathbb{R}} P^* \left[ \sqrt{T} \left( \hat{\theta}_T(\hat{\lambda}_{T,S,J,N}) - \hat{\theta}_T(\hat{\lambda}_{T,S,J,N}) \right) \right] \leq x \right) - P \left( \sqrt{T} \left( \hat{\theta}_T(\lambda^*) - \theta^*(\lambda^*) \right) \leq x \right) > \varepsilon
\]

\[
\to 0.
\]

By the definition of \( \hat{\theta}_T^* \left( \hat{\lambda}_{T,S,J,N} \right) \),

\[
0 = \nabla_{\theta} l^*_T(\hat{\theta}_T, \hat{\lambda}_{T,S,J,N}) + \nabla_{\theta \theta} l^*_T(\hat{\theta}_T, \hat{\lambda}_{T,S,J,N}) \left[ \hat{\theta}_T^*(\hat{\lambda}_{T,S,J,N}) - \hat{\theta}_T(\hat{\lambda}_{T,S,J,N}) \right],
\]

where \( \hat{\theta}_T^* \in \left( \hat{\theta}_T^*, \hat{\theta}_T \right) \). Thus,

\[
\sqrt{T} \left[ \hat{\theta}_T^*(\hat{\lambda}_{T,S,J,N}) - \hat{\theta}_T(\hat{\lambda}_{T,S,J,N}) \right] = \left[ -\nabla_{\theta \theta} l^*_T(\hat{\theta}_T, \hat{\lambda}_{T,S,J,N}) \right]^{-1} \sqrt{T} \nabla_{\theta} l^*_T(\hat{\theta}_T, \hat{\lambda}_{T,S,J,N})
\]

Now,

\[
E^* \left[ \sqrt{T} \nabla_{\theta} l^*_T(\hat{\theta}_T, \hat{\lambda}_{T,S,J,N}) \right] = \sqrt{T} \nabla_{\theta} l_T(\hat{\theta}_T, \hat{\lambda}_{T,S,J,N}) + O_P \left( \frac{1}{\sqrt{T}} \right) = o_P(1)
\]

given that \( \sqrt{T} \nabla_{\theta} l_T(\hat{\theta}_T, \hat{\lambda}_{T,S,J,N}) = 0 \) by the first order conditions, and \( l/T^{1/2} \to 0 \). Since each
block is independent of the others,

\[
\text{var}^* \left[ \sqrt{T} \nabla \theta^* l_T^i(\hat{\theta}_T, \hat{\lambda}_{T,S,I,N}) \right] = \text{var}^* \left[ \frac{1}{\sqrt{T}} \sum_{k=1}^{l_T} \sum_{i=1}^{l_T} \nabla \theta^* l_{K+i}^l(\hat{\theta}_T, \hat{\lambda}_{T,S,I,N}) \right]
\]

\[
= \text{E}^* \left[ \frac{1}{T} \sum_{k=1}^{b} \sum_{i=1}^{l_T} \sum_{j=1}^{l_T} \nabla \theta^* l_{K+i}^l(\hat{\theta}_T, \hat{\lambda}_{T,S,I,N}) \nabla \theta^* l_{K+j}^l(\hat{\theta}_T, \hat{\lambda}_{T,S,I,N})^\top \right]
\]

\[
= \frac{1}{T - l_T + 1} \frac{1}{l_T} \sum_{i=1}^{l_T} \sum_{j=1}^{l_T} \text{Pr}_P \left[ - \nabla^2 \theta^* l_i(\hat{\theta}_T, \hat{\lambda}_{T,S,I,N}) \nabla^2 \theta^* l_j(\hat{\theta}_T, \hat{\lambda}_{T,S,I,N})^\top \right]
\]

by Theorem 3.5 in Kunsch (1989). Finally,

\[
\left[ - \nabla^2 \theta^* l_i(\hat{\theta}_T, \hat{\lambda}_{T,S,I,N}) \right]^{-1} - \text{E} \left[ - \nabla^2 \theta^* l_i(\theta^1, \chi^1) \right] = o_P^*(1) \text{ Pr}_P
\]

where with the notation \(o_P^*(1) \text{ Pr}_P\) we mean a term approaching zero in the probability law of the bootstrap, conditionally on the sample and for all sample but a set of measure approaching zero.

**Proof of Lemma A3:** Straightforward from Lemma A2 and Theorem 3.

**Proof of Theorem 3:** For notation brevity, let

\[
G_{T,S,I,N}^{*,C} \left( \hat{\theta}_T^* (\lambda), \lambda \right) = G_{T,S,I,N}^* \left( \hat{\theta}_T^* (\lambda), \lambda \right) - G_{T,S,I,N} \left( \hat{\theta}_T (\lambda), \lambda \right)
\]

Via a mean value expansion around \(\hat{\lambda}_{T,S,I,N}\),

\[
\frac{1}{\sqrt{T}} \left( \hat{\lambda}_{T,S,I,N}^* - \hat{\lambda}_{T,S,I,N} \right)
\]

\[
= \left[ \nabla \lambda G_{T,S,I,N}^{*,C} \left( \hat{\theta}_T^* (\hat{\lambda}_{T,S,I,N}), \hat{\lambda}_{T,S,I,N}^* \right) V_{T,B} \left( \hat{\theta}_T (\hat{\lambda}_{T,S,I,N}), \hat{\lambda}_{T,S,I,N} \right)^{-1} \nabla \lambda G_{T,S,I,N}^{*,C} \left( \hat{\theta}_T (\hat{\lambda}_{T,S,I,N}), \hat{\lambda}_{T,S,I,N}^* \right) \right]
\]

\[
\times \frac{1}{\sqrt{T}} \left[ \nabla \lambda G_{T,S,I,N}^{*,C} \left( \hat{\theta}_T (\hat{\lambda}_{T,S,I,N}), \hat{\lambda}_{T,S,I,N}^* \right) V_{T,B} \left( \hat{\theta}_T (\hat{\lambda}_{T,S,I,N}), \hat{\lambda}_{T,S,I,N}^* \right)^{-1} \nabla \lambda G_{T,S,I,N}^{*,C} \left( \hat{\theta}_T (\hat{\lambda}_{T,S,I,N}), \hat{\lambda}_{T,S,I,N} \right) \right]
\]

\[
\times G_{T,S,I,N}^{*,C} \left( \hat{\theta}_T (\hat{\lambda}_{T,S,I,N}), \hat{\lambda}_{T,S,I,N} \right)
\]

27
Now, by the same argument used in the proof of Theorem 3, \( \sqrt{T}G_{T,S,J,N}^{\ast,e} \left( \hat{\theta}_T^{\ast} (\hat{\lambda}_{T,S,J,N}^{\ast}), \hat{\lambda}_{T,S,J,N}^{\ast} \right) \) has the same limiting distribution as \( \sqrt{T}G_{T,S,J,N} \left( \hat{\theta}_T (\lambda^{\ast}), \lambda^{\ast} \right) \). Furthermore,

\[
\nabla_{\lambda} G_{T,S,J,N}^{\ast,e} \left( \hat{\theta}_T^{\ast} (\hat{\lambda}_{T,S,J,N}^{\ast}), \hat{\lambda}_{T,S,J,N}^{\ast} \right) - A_{\lambda}^{\dagger} = o_{P}^{\ast}(1) \quad \text{Pr}-P,
\]

where

\[
A_{\lambda}^{\dagger} = \text{plim}_{T,S,J \to \infty} \nabla_{\lambda} G_{T,S,J,N} \left( \theta^{\dagger}(\lambda^{\ast}), \lambda^{\ast} \right).
\]

The statement then follows from Theorem 3.
References


Table 1
Moment conditions

<table>
<thead>
<tr>
<th>Data Model</th>
<th>Data</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Returns</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.3977</td>
<td>2.4373</td>
<td></td>
</tr>
<tr>
<td>( a^R )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-1.4412)</td>
<td>(-1.0090)</td>
<td></td>
</tr>
<tr>
<td>( b_{1,12}^R )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-0.8335)</td>
<td>(-1.3512)</td>
<td></td>
</tr>
<tr>
<td>( b_{2,12}^R )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volatility</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-0.3982)</td>
<td>0.0103</td>
<td></td>
</tr>
<tr>
<td>( a^V )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0136</td>
<td>0.8732</td>
<td></td>
</tr>
<tr>
<td>( \phi_6 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-0.6863)</td>
<td>(-0.5546)</td>
<td></td>
</tr>
<tr>
<td>( \phi_{12} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5289</td>
<td>0.4313</td>
<td></td>
</tr>
<tr>
<td>( \phi_{18} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-0.3225)</td>
<td>(-0.1561)</td>
<td></td>
</tr>
<tr>
<td>( \phi_{24} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0572</td>
<td>0.0327</td>
<td></td>
</tr>
<tr>
<td>( \phi_{36} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-0.0195)</td>
<td>0.0169</td>
<td></td>
</tr>
<tr>
<td>( \phi_{48} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0709</td>
<td>0.1574</td>
<td></td>
</tr>
<tr>
<td>( b_{1,12}^V )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-0.1090)</td>
<td>(-0.0017)</td>
<td></td>
</tr>
<tr>
<td>( b_{1,24}^V )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1769</td>
<td>(-0.0292)</td>
<td></td>
</tr>
<tr>
<td>( b_{1,36}^V )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0064</td>
<td>(-0.0642)</td>
<td></td>
</tr>
<tr>
<td>( b_{1,48}^V )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0458</td>
<td>(-0.0352)</td>
<td></td>
</tr>
<tr>
<td>( b_{2,12}^V )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0967</td>
<td>0.0206</td>
<td></td>
</tr>
<tr>
<td>( b_{2,24}^V )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1082</td>
<td>(-0.0156)</td>
<td></td>
</tr>
<tr>
<td>( b_{2,36}^V )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0356</td>
<td>0.0078</td>
<td></td>
</tr>
<tr>
<td>( b_{2,48}^V )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \mu_S )</td>
<td>2.8842</td>
<td>2.9002</td>
</tr>
<tr>
<td>( V_S )</td>
<td>0.1136</td>
<td>0.1475</td>
</tr>
<tr>
<td>( VIX )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.4585</td>
<td>2.6766</td>
<td></td>
</tr>
<tr>
<td>( a^{VIX} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-2.2466)</td>
<td>(-0.1324)</td>
<td></td>
</tr>
<tr>
<td>( b_{1,1}^{VIX} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-1.8851)</td>
<td>0.0513</td>
<td></td>
</tr>
<tr>
<td>( b_{1,12}^{VIX} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-0.1323)</td>
<td>(-0.0501)</td>
<td></td>
</tr>
<tr>
<td>( b_{1,24}^{VIX} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-0.5017)</td>
<td>(-0.8389)</td>
<td></td>
</tr>
<tr>
<td>( b_{1,36}^{VIX} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-0.4266)</td>
<td>(-1.4212)</td>
<td></td>
</tr>
<tr>
<td>( b_{1,48}^{VIX} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-0.3072)</td>
<td>(-0.0073)</td>
<td></td>
</tr>
<tr>
<td>( b_{2,1}^{VIX} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4095</td>
<td>(-0.1108)</td>
<td></td>
</tr>
<tr>
<td>( b_{2,24}^{VIX} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9330</td>
<td>0.1113</td>
<td></td>
</tr>
<tr>
<td>( b_{2,36}^{VIX} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \mu^{VIX} )</td>
<td>0.1894</td>
<td>0.2018</td>
</tr>
<tr>
<td>( \sigma^{VIX} )</td>
<td>0.0636</td>
<td>0.0555</td>
</tr>
</tbody>
</table>
### Table 2
Parameter estimates for the macroeconomic factors

<table>
<thead>
<tr>
<th>Estimate</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa_1$</td>
<td>0.0255</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>1.0379</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.0059</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>−0.0054</td>
</tr>
<tr>
<td>$\kappa_2$</td>
<td>0.5628</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>1.0388</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.0517</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>−0.0479</td>
</tr>
<tr>
<td>$\bar{\kappa}_1$</td>
<td>−0.2532</td>
</tr>
<tr>
<td>$\bar{\kappa}_2$</td>
<td>1.1701</td>
</tr>
</tbody>
</table>

### Table 3
Parameter estimates for the stock price process and the unobservable factor

<table>
<thead>
<tr>
<th>Estimate</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_0$</td>
<td>0.2356</td>
</tr>
<tr>
<td>$s_1$</td>
<td>0.0629</td>
</tr>
<tr>
<td>$s_2$</td>
<td>2.4437</td>
</tr>
<tr>
<td>$s_3$</td>
<td>0.2371</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>2.3022</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.2055</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>0.0493</td>
</tr>
<tr>
<td>$\kappa_3$</td>
<td>1.2097·10$^{-2}$</td>
</tr>
</tbody>
</table>

### Table 4
Parameter estimates for the risk-premium process

<table>
<thead>
<tr>
<th>Estimate</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_{1(1)}$</td>
<td>1.2341·10$^{-1}$</td>
</tr>
<tr>
<td>$\lambda_{2(1)}$</td>
<td>−1.0021</td>
</tr>
<tr>
<td>$\lambda_{1(2)}$</td>
<td>1.0214·10$^3$</td>
</tr>
<tr>
<td>$\lambda_{2(2)}$</td>
<td>1.4594·10$^{-1}$</td>
</tr>
<tr>
<td>$\lambda_{1(3)}$</td>
<td>5.5424·10$^{-2}$</td>
</tr>
<tr>
<td>$\lambda_{2(3)}$</td>
<td>1.0024</td>
</tr>
</tbody>
</table>
Figure 1 – Industrial production growth and inflation, with NBER dated recession periods. This figure plots the one-year, monthly gross inflation, defined as $y_{1,t} \equiv \frac{CPI_t}{CPI_{t-12}}$, and the one-year, monthly gross industrial production growth, defined as $y_{2,t} \equiv \frac{IP_t}{IP_{t-12}}$, where CPI$_t$ is the Consumer price index as of month $t$, and IP$_t$ is the real, seasonally adjusted industrial production index as of month $t$. The sample covers monthly data for the period from January 1950 to December 2006. Vertical solid lines (in black) track the beginning of NBER-dated recessions, and vertical dashed lines (in red) indicate the end of NBER-dated recessions.
Figure 2 – Returns and volatility along with the model predictions, with NBER dated recession periods. This figure plots one-year ex-post price changes and one-year return volatility, along with their counterparts predicted by the model. The top panel depicts continuously compounded price changes, defined as $R_t \equiv \log \left( \frac{s_t}{s_{t-12}} \right)$, where $s_t$ is the real stock price as of month $t$. The middle panel depicts smoothed return volatility, defined as $\text{Vol}_t \equiv \sqrt{\frac{6}{\pi}} \cdot 12^{-1} \sum_{i=1}^{12} |\log \left( \frac{s_{t+1-i}}{s_{t-1}} \right)|$. The bottom panel plots the instantaneous standard deviation predicted by the model, obtained through Eq. (4), and the smoothed measured volatility, $\text{Vol}_t$. Each prediction at each point in time is obtained by feeding the model with the two macroeconomic factors depicted in Figure 1 (inflation and growth) and by averaging over 1000 dynamic simulations of the unobserved factor. The sample covers monthly data for the period from January 1950 to December 2006. Vertical solid lines (in black) track the beginning of NBER-dated recessions, and vertical dashed lines (in red) indicate the end of NBER-dated recessions.
Figure 3 — VIX Index and the volatility risk-premium along with model predictions, with NBER dated recession periods. This figure plots the VIX index along with model’s predictions. The top panel depicts the VIX index and (i) the VIX index predicted by the model; (ii) the VIX index predicted by the model in an economy without risk-aversion, i.e. the expected integrated volatility under the physical probability; and (iii) the instantaneous volatility predicted by the model, obtained through Eq. (4). The bottom panel depicts the volatility risk-premium predicted by the model, defined as the difference between the model-generated expected integrated volatility under the risk-neutral and physical probability,

$$\text{VRP} (\mathbf{y}(t)) \equiv \sqrt{\frac{1}{T-t}} \left( \sqrt{E_Q \left( \int_t^T \sigma^2 (\mathbf{y}(u)) \, du \bigg| \mathbf{y}(t) \right)} - \sqrt{E \left( \int_t^T \sigma^2 (\mathbf{y}(u)) \, du \bigg| \mathbf{y}(t) \right)} \right),$$

where $T - t = 12^{-1}$, $E_Q$ is the conditional expectation under the risk-neutral probability, $\sigma^2 (\mathbf{y})$ is the instantaneous variance predicted by the model, obtained through Eq. (4), and $\mathbf{y}$ is the vector of three factors: the two macroeconomic factors depicted in Figure 1 (inflation and growth) and one unobservable factor. Each prediction at each point in time is obtained by feeding the model with the two macroeconomic factors depicted in Figure 1 (inflation and growth) and by averaging over 1000 dynamic simulations of the unobserved factor. The sample covers monthly data for the period from January 1990 to December 2006. Vertical solid lines (in black) track the beginning of NBER-dated recessions, and vertical dashed lines (in red) indicate the end of NBER-dated recessions.
Figure 4 — Volatility risk-premium against inflation and industrial production growth. This figure provides scatterplots of the volatility risk-premium predicted by the model depicted in Figure 3 (bottom panel) against the two macroeconomic factors depicted in Figure 1 (inflation and growth). Each prediction at each point in time is obtained by feeding the model with the two macroeconomic and by averaging over 1000 dynamic simulations of the unobserved factor. The sample covers monthly data for the period from January 1990 to December 2006. Vertical solid lines (in black) track the beginning of NBER-dated recessions, and vertical dashed lines (in red) indicate the end of NBER-dated recessions.