Macroeconomic Determinants of Stock Market Volatility and Volatility Risk-Premia*

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

First draft: July 22, 2005
This draft: March 3, 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 is 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 and that in turn, they are of substantial help in predicting future economic activity.

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

* *macro_vol.rap*. We thank Christian Julliard, Mark Salmon and seminar participants at 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. The usual disclaimer applies. This version of the paper is also incomplete as regards the review of the literature on inference methods for nonlinear asset pricing models.

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 such as investment bankers 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 because it accounts for no-arbitrage information. Finally, we use the model to produce joint forecasts of both economic activity and stock market volatility.

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 related 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.

[Give details on our econometric methodology here]

The remainder of the paper is organized as follows. In Section 2 we develop the no-arbitrage model for stock price, volatility and variance risk-premia. Section 3 contains the estimation strategy. Section 4 presents our empirical results, and an 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 $\boldsymbol{y}(t)$, solution to a n-dimensional affine diffusion,

$$\mathrm{d}\boldsymbol{y}(t) = \boldsymbol{\kappa}(\boldsymbol{\mu} - \boldsymbol{y}(t))\,\mathrm{d}t + \boldsymbol{\Sigma}\boldsymbol{V}(\boldsymbol{y}(t))\,\mathrm{d}\boldsymbol{W}(t), \tag{1}$$

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

$$\boldsymbol{V}(\boldsymbol{y})_{(ii)} = \sqrt{\alpha_i + \boldsymbol{\beta}_i^\top \boldsymbol{y}}, \quad i = 1, \cdots, d,$$

for some scalars α_i and vectors $\boldsymbol{\beta}_i$. Appendix A reviews sufficient conditions that are known to ensure that Eq. (1) has a strong solution with $\boldsymbol{V}(\boldsymbol{y}(t))_{(ii)} > 0$ almost surely for all t.

While we do not necessarily observe every single component of $\boldsymbol{y}(t)$, we do observe discretely sampled paths of macroeconomic variables such as industrial production, unemployment or inflation. Let $\{\mathrm{M}_j(t)\}_{t=1,2,\cdots}$ be the discretely sampled path of the macroeconomic variable $\mathrm{M}_j(t)$ where, for example, $\mathrm{M}_j(t)$ can be the industrial production index available at time t, and $j = 1, \cdots, N_{\mathrm{M}}$, 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:

$$\log(\mathrm{M}_j(t)/\mathrm{M}_j(t-12)) = \varphi_j(\boldsymbol{y}(t)), \quad j = 1, \cdots, N_{\mathrm{M}}, \tag{2}$$

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 $\boldsymbol{y}(t)$ 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 $s(\boldsymbol{y}(t))$ such that the real stock price at time t, $s(t)$ say, is $s(t) \equiv s(\boldsymbol{y}(t))$. We let this price function be twice continuously differentiable in $\boldsymbol{y}$. (Given the assumptions and conditions we give below, this differentiability condition holds in our model.) By Itô's lemma, $s(t)$ satisfies,

$$\frac{\mathrm{d}s(t)}{s(t)} = m(\boldsymbol{y}(t), s(t))\,\mathrm{d}t + \frac{s_y(\boldsymbol{y}(t))^\top \boldsymbol{\Sigma} \boldsymbol{V}(\boldsymbol{y}(t))}{s(\boldsymbol{y}(t))} \mathrm{d}\boldsymbol{W}(t), \tag{3}$$

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

$$\sigma(t)^2 \equiv \left\| \frac{s_y(\boldsymbol{y}(t))^\top \boldsymbol{\Sigma} \boldsymbol{V}(\boldsymbol{y}(t))}{s(\boldsymbol{y}(t))} \right\|^2. \tag{4}$$

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,

$$\xi(T) \equiv \frac{\mathrm{d}Q}{\mathrm{d}P} = \exp\left(-\int_0^T \boldsymbol{\Lambda}(t)^\top \mathrm{d}\boldsymbol{W}(t) - \frac{1}{2}\int_0^T \|\boldsymbol{\Lambda}(t)\|^2 \mathrm{d}t \right),$$

for some adapted $\boldsymbol{\Lambda}(t)$, the risk-premium process. We assume that each component of the risk-premium *process* $\Lambda^i(t)$ satisfies,

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

for some *function* Λ^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,

$$\frac{\mathrm{d}s(t)}{s(t)} = (r - \delta(\boldsymbol{y}(t)))\,\mathrm{d}t + \frac{s_y(\boldsymbol{y}(t))^\top \boldsymbol{\Sigma} \boldsymbol{V}(\boldsymbol{y}(t))}{s(\boldsymbol{y}(t))} \mathrm{d}\hat{\boldsymbol{W}}(t), \tag{5}$$

where $\delta(\boldsymbol{y})$ is the instantaneous dividend rate, and $\hat{\boldsymbol{W}}$ 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 $\boldsymbol{V}^{-}(\boldsymbol{y})$ be a $d \times d$ diagonal matrix with elements

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

and set,

$$\boldsymbol{\Lambda}(\boldsymbol{y}) = \boldsymbol{V}(\boldsymbol{y})\boldsymbol{\lambda}_1 + \boldsymbol{V}^{-}(\boldsymbol{y})\boldsymbol{\lambda}_2\boldsymbol{y}, \tag{6}$$

for some d-dimensional vector $\boldsymbol{\lambda}_1$ and some $d \times n$ matrix $\boldsymbol{\lambda}_2$. The functional form for $\boldsymbol{\Lambda}$ is the same as in the specification suggested by Duffee (2002) in the term-structure literature. If the matrix $\boldsymbol{\lambda}_2 = \boldsymbol{0}_{d\times n}$, then, $\boldsymbol{\Lambda}$ collapses to the standard "completely affine" specification introduced by Duffie and Kan (1996), in which the risk-premia $\boldsymbol{\Lambda}$ are tied up to the volatility of the fundamentals, $\boldsymbol{V}(\boldsymbol{y})$. 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 $\boldsymbol{\lambda}_2\boldsymbol{y}$.

Finally, we determine the no-arbitrage stock price. Let us assume that for all $\boldsymbol{y}$, $s_y(\boldsymbol{y})^\top \boldsymbol{\Sigma}\boldsymbol{V}(\boldsymbol{y})$ satisfies some regularity conditions (local martingale $\to$ martingale). Assuming no-bubbles, Eq. (5) implies that the stock price is,

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

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 $\boldsymbol{y}$. As it turns out, the previous assumption on the pricing kernel and the assumption that $\delta(\cdot)$ is affine in $\boldsymbol{y}$ implies that the stock price is also affine in $\boldsymbol{y}$. Precisely, let

$$\delta(\boldsymbol{y}) = \delta_0 + \boldsymbol{\delta}^\top\boldsymbol{y}, \tag{8}$$

for some scalar δ_0 and some vector $\boldsymbol{\delta}$. 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 $s(\boldsymbol{y})$ is linear in the state vector $\boldsymbol{y}$, viz*

$$s(\boldsymbol{y}) = \frac{\delta_0 + \boldsymbol{\delta}^\top(\boldsymbol{D} + r\boldsymbol{I}_{n\times n})^{-1}\boldsymbol{c}}{r} + \boldsymbol{\delta}^\top(\boldsymbol{D} + r\boldsymbol{I}_{n\times n})^{-1}\boldsymbol{y}, \tag{9}$$

where

$$\begin{aligned} \boldsymbol{c} &= \boldsymbol{\kappa}\boldsymbol{\mu} - \boldsymbol{\Sigma}\begin{pmatrix} \alpha_1\lambda_{1(1)} & \cdots & \alpha_d\lambda_{1(d)} \end{pmatrix}^\top, \\ \boldsymbol{D} &= \boldsymbol{\kappa} + \boldsymbol{\Sigma}\left[\begin{pmatrix} \lambda_{1(1)}\boldsymbol{\beta}_1^\top & \cdots & \lambda_{1(d)}\boldsymbol{\beta}_d^\top \end{pmatrix}^\top + \boldsymbol{I}^{-}\boldsymbol{\lambda}_2\right], \end{aligned}$$

$\boldsymbol{I}^{-}$ is a $d \times d$ diagonal matrix with elements $\boldsymbol{I}^{-}_{(ii)} = 1$ if $\Pr\{\boldsymbol{V}(\boldsymbol{y}(t))_{(ii)} > 0$ all $t\} = 1$ and 0 otherwise; and, finally $\{\lambda_{1(j)}\}_{j=1}^{d}$ are the components of $\boldsymbol{\lambda}_1$.

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,

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

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, 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,

$$\begin{cases} \dfrac{\mathrm{d}s(t)}{s(t)} = m(t)\,\mathrm{d}t + v(t)\,\mathrm{d}W_1(t) \\ \mathrm{d}v(t)^2 = \kappa\left(\mu - v(t)^2\right)\mathrm{d}t + \sigma v(t)\left(\rho \mathrm{d}W_1(t) + \sqrt{1-\rho^2}\mathrm{d}W_2(t)\right) \end{cases} \tag{11}$$

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.[1] 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 volatility on a given interval $[t, T]$ as,

$$IV_{t,T} = \int_t^T \left(\frac{\mathrm{d}}{\mathrm{d}\tau} var \left[\log s(\tau) \middle| \mathbb{F}(u) \right] \Big|_{\tau=u} \right) \mathrm{d}u. \tag{12}$$

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 of the future integrated volatility is a functional of put and call options written on the asset:

$$\mathbb{E}\left[IV_{t,T} \middle| \mathbb{F}(t) \right] = 2 \left[\int_0^{F(t)} \frac{P(t,T,K)}{u(t,T)} \frac{1}{K^2} \mathrm{d}K + \int_{F(t)}^{\infty} \frac{C(t,T,K)}{u(t,T)} \frac{1}{K^2} \mathrm{d}K \right], \tag{13}$$

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 integrated volatility. Our model predicts that the risk-neutral expectation of integrated volatility is:

$$\mathbb{E}\left[IV_{t,T} \middle| \mathbb{F}(t) \right] = \int_t^T \mathbb{E}[\sigma(t)^2 \mid \mathbb{F}(t)] \mathrm{d}u,$$

where $\mathbb{F}(t)$ 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 indexe that matches that computed through Eq. (13).

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

$$VR(t) = E\left[IV_{t,T} \middle| \mathbb{F}(t) \right] - \mathbb{E}\left[IV_{t,T} \middle| \mathbb{F}(t) \right], \tag{14}$$

changes with changes in the factors $\boldsymbol{y}(t)$ 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

[1] If the interest rate is zero, then, in the absence of arbitrage opportunities, the variance swap rate is simply the expectation of future integrated return volatility under the risk-neutral probability, as defined in Eq. (12) below.

the only observable factors (say y_1 and y_2) affecting the stock market development. We define these factors as follows:

$$\log\left(\mathrm{M}_j(t)/\,\mathrm{M}_j(t-12)\right) = \log y_j(t), \quad j = 1, 2,$$

where $\mathrm{M}_1(t)$ is the consumer price index as of month t and $\mathrm{M}_2(t)$ 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 $\varphi_j(\boldsymbol{y}) = \log y_j$.

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),

$$\boldsymbol{\kappa} = \begin{bmatrix} \kappa_1 & \bar{\kappa}_1 & 0 \\ \bar{\kappa}_2 & \kappa_2 & 0 \\ 0 & 0 & \kappa_3 \end{bmatrix},$$

where κ_1 and κ_2 are the speed of adjustment of inflation and industrial production growth towards their long-run means μ_1 and μ_2, in the absence of feedbacks, and $\bar{\kappa}_1$ and $\bar{\kappa}_2$ are the feedback parameters. Moreover, we take $\boldsymbol{\Sigma} = \boldsymbol{I}_{3\times 3}$ and the vectors $\boldsymbol{\beta}_i$ so as to make y_j solution to,

$$\mathrm{d}y_j(t) = \left[\kappa_j\left(\mu_j - y_j(t)\right) + \bar{\kappa}_j\left(\bar{\mu}_j - \bar{y}_j(t)\right)\right]\mathrm{d}t + \sqrt{\alpha_j + \beta_j y_j(t)}\,\mathrm{d}W_j(t), \quad j = 1, 2, 3, \qquad (15)$$

where, for brevity, we have set $\bar{\mu}_1 \equiv \mu_2$, $\bar{y}_1(t) \equiv y_2(t)$, $\bar{\mu}_2 \equiv \mu_1$, $\bar{y}_2(t) = y_1(t)$, $\bar{\kappa}_3 \equiv \bar{\mu}_3 \equiv \bar{y}_3(t) \equiv 0$ and, finally, $\beta_j \equiv \beta_{jj}$. We assume that $\Pr\{\boldsymbol{V}(\boldsymbol{y}(t))_{(ii)} > 0 \text{ all } t\} = 1$, which it does under the conditions reviewed in Appendix A.

We assume that the risk-premium process $\boldsymbol{\Lambda}$ satisfies the "essentially affine" specification in Eq. (6), where we take the matrix $\boldsymbol{\lambda}_2$ 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,

$$\boldsymbol{\pi}(\boldsymbol{y}) \equiv \boldsymbol{\Sigma}\boldsymbol{V}(\boldsymbol{y})\boldsymbol{\Lambda}(\boldsymbol{y}) = \begin{pmatrix} \alpha_1\lambda_{1(1)} + \left(\beta_1\lambda_{1(1)} + \lambda_{2(1)}\right)y_1 \\ \alpha_2\lambda_{1(2)} + \left(\beta_2\lambda_{1(2)} + \lambda_{2(2)}\right)y_2 \\ \alpha_3\lambda_{1(3)} + \left(\beta_3\lambda_{1(3)} + \lambda_{2(3)}\right)y_3 \end{pmatrix} \qquad (16)$$

depends on the factor y_j not only through the channel of the volatility of these factors (i.e. through the parameters β_{jj}), but also through the additional risk-premia parameters $\lambda_{2(j)}$.

Finally, the instantaneous dividend process $\delta(t)$ in (8) satisfies,

$$\delta(\boldsymbol{y}) = \delta_0 + \delta_1 y_1 + \delta_2 y_2 + \delta_3 y_3. \qquad (17)$$

Under these conditions, the asset price in Proposition 1 is given by,

$$s(\boldsymbol{y}) = s_0 + \sum_{j=1}^{3} s_j y_j, \qquad (18)$$

where

$$s_0 = \frac{1}{r}\left[\delta_0 + \sum_{j=1}^{3} s_j\left(\kappa_j\mu_j + \bar{\kappa}_j\bar{\mu}_j - \alpha_j\lambda_{1(j)}\right)\right], \tag{19}$$

$$s_j = \frac{\delta_j\left(r+\kappa_j-\bar{\kappa}_j+\beta_j\lambda_{1(j)}+\lambda_{2(j)}\right)}{\prod_{h=1}^{2}\left(r+\kappa_h+\beta_h\lambda_{1(h)}+\lambda_{2(h)}\right)-\bar{\kappa}_1\bar{\kappa}_2\left(1-\mathbb{I}_{(j=3)}\right)}, \quad j=1,2,3, \tag{20}$$

and where $\bar{\kappa}_j$ and $\bar{\mu}_j$ as 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 $\lambda_{(2)i}$ and δ_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 $\boldsymbol{y}$. Instead, the parameters δ_i determine how sensitive the dividend process in Eq. (17) is to changes in $\boldsymbol{y}$. 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 $\boldsymbol{\lambda} = \left(\left(\lambda_{1(j)}, \lambda_{2(j)}\right)_{j=1,2,3}\right)$ and $\boldsymbol{\theta} = \left(\delta_0, \left(\kappa_j, \mu_j, \alpha_j, \beta_j, \delta_j\right)_{j=1,2,3}, \left(\bar{\kappa}_j\right)_{j=1,2}\right)$. 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 $\boldsymbol{\lambda}$ as a vector of unidentified nuisance parameters, and estimate $\boldsymbol{\theta}$ by maximum likelihood, for any given value of $\boldsymbol{\lambda}$. In the second step, we use the Arrow-Debreu adjusted volatility in Section 2.4 and calibrate $\boldsymbol{\lambda}$ to reconcile the model's predictions with the data.

3.1 Estimation of $\boldsymbol{\theta}$, for given $\boldsymbol{\lambda}$

Let $\mathbf{y}_t = (y_{1,t}, y_{2,t}, y_{3,t})$ denote the skeleton of $\mathbf{y}(t)$ in Eq. (2), sampled at $t = 1, \cdots, T$, and set $\mathbf{x}_t = (y_{1,t}, y_{2,t})$. Although the conditional density of $\mathbf{x}_{t+1}$ given $\mathbf{x}_t$ is unknown in closed-form,[2] we may rely on a variety of approaches to approximate this density. For example, we might rely on the work of Duffie, Pan and Singleton (2000) and Singleton (2001) to accomplish this task. In this paper, we use Aït-Sahalia [**????**] small time expansions to obtain parameter estimates for the observable variables, i.e. estimates of $\boldsymbol{\theta}_{\mathrm{O}} = \left(\left(\kappa_j, \mu_j, \alpha_j, \beta_j\right)_{j=1,2}, \left(\bar{\kappa}_j\right)_{j=1,2}\right)$. The maximum likelihood estimator $\hat{\boldsymbol{\theta}}_{\mathrm{O},T}$ is,

$$\hat{\boldsymbol{\theta}}_{\mathrm{O},T} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} l_T\left(\boldsymbol{\theta}_{\mathrm{O}}\right)$$

[2] Unless $\bar{\kappa}_i = 0$ and, either $\alpha_i = 0$ or $\beta_i = 0$, for $i = 1, 2, 3$.

where,

$$l_T(\boldsymbol{\theta}_{\mathrm{O}}) = \frac{1}{T}\sum_{t=1}^{T} l_t(\mathbf{x}_{t+1}|\mathbf{x}_t;\boldsymbol{\theta}_{\mathrm{O}}); \quad l_t(\mathbf{x}_{t+1}|\mathbf{x}_t;\boldsymbol{\theta}_{\mathrm{O}}) = \log\varphi_K(\mathbf{x}_{t+1}|\,\mathbf{x}_t;\boldsymbol{\theta}_{\mathrm{O}}), \tag{21}$$

and $\varphi_K(\mathbf{x}_t|\,\mathbf{x}_{t-1};\boldsymbol{\theta}_{\mathrm{O}})$ is the conditional density obtained through the Aït-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 $\delta(\boldsymbol{y})$ (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 $\boldsymbol{\lambda}$, it is observationally equivalent to estimate $\boldsymbol{\theta}_{\mathrm{U}} = \left(s_0, (s_j)_{j=1,2,3}, \kappa_3, \mu_3, \alpha_3, \beta_3\right)$ or $\left(\delta_0, (\delta_j)_{j=1,2,3}, \kappa_3, \mu_3, \alpha_3, \beta_3\right)$. We estimate $\boldsymbol{\theta}_{\mathrm{U}}$, by creating moment conditions. We want to calibrate $\boldsymbol{\theta}_{\mathrm{U}}$ in such a way that the model predicts patterns of ex-post returns and ex-post return volatility that match their empirical counterparts, described in the empirical section. For any given $\boldsymbol{\lambda}$, define,

$$F_{T,J}(\boldsymbol{\theta}_{\mathrm{U}}(\boldsymbol{\lambda}), \boldsymbol{\lambda}) = \frac{1}{J}\sum_{j=1}^{J} f_j(\boldsymbol{\theta}_{\mathrm{U}}(\boldsymbol{\lambda}), \boldsymbol{\lambda}) - \frac{1}{T}\sum_{t=1}^{T} f_t,$$

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

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

We need the following assumptions:

Assumptions:
A1: $\mathrm{E}\nabla_{\boldsymbol{\theta}}\left(l_T(\boldsymbol{\theta}^{\dagger}_{\mathrm{U}}(\boldsymbol{\lambda}), \boldsymbol{\lambda})\right) > \mathrm{E}\nabla_{\boldsymbol{\theta}}\left(l_T(\boldsymbol{\theta}_U(\boldsymbol{\lambda}), \boldsymbol{\lambda})\right)$ for all $\boldsymbol{\theta} \neq \boldsymbol{\theta}^{\dagger}$, for all $\boldsymbol{\lambda}$
A2: $\mathrm{E}\left(-\nabla^2_{\boldsymbol{\theta}}()???\right)$ uniformly positive definite in $\boldsymbol{\theta}\times\boldsymbol{\lambda}$
A3: $\mathrm{E}\left[(\nabla_{\boldsymbol{\theta}} l??????))^{4+\delta}\right]_i < \infty$, $i = 1, \cdots, k$, for some $\delta > 0$, where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ with $\boldsymbol{\Theta}$ being a compact subset of $\mathbb{R}^k$.

We have:

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

(i)

$$\sup_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}}\left|\hat{\boldsymbol{\theta}}_{\mathrm{U},T}(\boldsymbol{\lambda}) - \boldsymbol{\theta}^{\dagger}_{\mathrm{U}}(\boldsymbol{\lambda})\right| = o_P(1).$$

(ii) Pointwise in $\boldsymbol{\lambda}$,

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_{\mathrm{U},T}(\boldsymbol{\lambda}) - \boldsymbol{\theta}^{\dagger}_{\mathrm{U}}(\boldsymbol{\lambda})\right) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \mathbf{???}),$$

where

$$\boldsymbol{B}(\boldsymbol{\theta}_{\mathrm{O}}^{\dagger}(\boldsymbol{\lambda}))^{-1} = -\mathrm{E}\left(\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} l_T(\boldsymbol{\theta}_{\mathrm{O}}^{\dagger}, \boldsymbol{\lambda})\right) \quad \textit{and} \quad \boldsymbol{A}(\boldsymbol{\theta}_{\mathrm{O}}^{\dagger}(\boldsymbol{\lambda})) = \lim_{T\to\infty} \mathrm{var}\left(\sqrt{T}\nabla_{\boldsymbol{\theta}} l_T(\boldsymbol{\theta}_{\mathrm{O}}^{\dagger}, \boldsymbol{\lambda})\right).$$

Note that we compute $\hat{\boldsymbol{\theta}}_T(\boldsymbol{\lambda}) = \left(\hat{\boldsymbol{\theta}}_{\mathrm{O},T}, \hat{\boldsymbol{\theta}}_{\mathrm{U},T}(\boldsymbol{\lambda})\right)$ 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 $\boldsymbol{\theta}$ 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 $\boldsymbol{\theta}$. We aim to construct parametric and nonparametric measures of the conditional expectation of integrated volatility, under the risk neutral probability measure, i.e. $\mathbb{E}\left[IV_{t,t+1}\middle|\mathcal{F}(t)\right]$. We identify $\boldsymbol{\lambda}$ 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 $\mathbb{E}\left[IV_{t,t+1}\middle|\mathcal{F}(t)\right]$. 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 $\boldsymbol{\Sigma} = \boldsymbol{I}_{3\times 3}$, this is given by

$$\mathrm{d}\boldsymbol{y}(t) = \left[\boldsymbol{\kappa}\left(\boldsymbol{\mu} - \boldsymbol{y}(t)\right) - \boldsymbol{V}\left(\boldsymbol{y}(t)\right)\boldsymbol{\Lambda}\left(\boldsymbol{y}(t)\right)\right]\mathrm{d}t + \boldsymbol{V}\left(\boldsymbol{y}(t)\right)\mathrm{d}\hat{\boldsymbol{W}}(t), \tag{22}$$

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

$$\begin{aligned} y_{\ell,(k+1)\xi}^{(i)} &= y_{\ell,k\xi}^{(i)} + \left[\boldsymbol{\kappa}(\boldsymbol{\mu} - \boldsymbol{y}_{k\xi}^{(i)}) - \boldsymbol{V}(\boldsymbol{y}_{k\xi}^{(i)})\boldsymbol{\Lambda}(\boldsymbol{y}_{k\xi}^{(i)})\right]_{\ell}\xi + \sqrt{\alpha_\ell + \beta_\ell y_{\ell,k\xi}^{(i)}}\,\epsilon_{\ell,(k+1)\xi}^{(i)} \\ &- \frac{1}{2}\left\|\frac{\partial}{\partial\theta}\sqrt{\alpha_\ell + \beta_\ell y_{i,k\xi}^{(i)}}\right\|^2 + \frac{1}{2}\left\|\frac{\partial}{\partial\theta}\sqrt{\alpha_\ell + \beta_\ell y_{\ell,k\xi}^{(i)}}\right\|^2\left(\epsilon_{\ell,(k+1)\xi}^{(i)}\right)^2, \end{aligned} \tag{23}$$

where ϵ is NID with variance ξ, and assuming without loss of generality that ξ^{-1} is an integer. For each simulation i, we compute,

$$\log s_{k,\xi}^{(i)} \equiv \log\left(s_0 + \sum_{\ell=1}^{3} s_\ell y_{\ell,k\xi}^{(i)}\right). \tag{24}$$

The processes in (23) and (24) are simulated for each value of $\boldsymbol{\lambda}$, with parameter $\boldsymbol{\theta}$ fixed at $\hat{\boldsymbol{\theta}}_T(\boldsymbol{\lambda})$. We fix J sets of initial values for the vector of factors, i.e. $\mathbf{y}^j(0)$, $j = 1, \cdots, J$. For each j, we simulate (22) and (24) using $\boldsymbol{\lambda}$ and $\hat{\boldsymbol{\theta}}_T(\boldsymbol{\lambda})$, and for each $\boldsymbol{\lambda}$ we simulate S paths of length one.[3] For each initial value $\mathbf{y}^j(0)$, $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}$, and for each simulation replication i, we thus obtain $\log s_{k\xi}^{(i)}(\hat{\boldsymbol{\theta}}_T(\boldsymbol{\lambda}), \boldsymbol{\lambda}, \mathbf{y}^j(0))$, for $k = 1, \cdots, \xi^{-1}$. For each $j = 1, \cdots, J$, define,

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

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

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

and for $S/T \to 0$,

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

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

Next, we turn to the model-free estimator of $\mathbb{E}\left[IV_{t,t+1}|\, \mathcal{F}(t)\right]$ 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),

$$\widehat{\mathbb{E}\left[IV_{t,t+1}|\, \mathbb{F}(t)\right]} = \frac{2}{N}\left[\sum_{i:K_i \le F(t)}^{N} \frac{P(t, t+1, K_i)}{u(t, t+1)}\frac{1}{K_i^2} + \sum_{i:K_i > F(t)}^{N} \frac{C(t, t+1, K_i)}{u(t, t+1)}\frac{1}{K_i^2}\right], \tag{25}$$

for $t = 1, \cdots, T$.[4] Finally, define

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

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

[3] In practice we disregard a sufficiently high number of draws. However, for notational simplicity, we proceed as if we were to use $\mathbf{y}^j(0)$ as the initial value.

[4] 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.

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 kth autocovariance of model implied and model free integrated volatility, under the risk neutral probability measure. As shown below, $\boldsymbol{\lambda}$ can be estimated by minimizing the distance between moment conditions based on parametric and nonparametric estimator.

3.3 A first step estimator of $\boldsymbol{\lambda}$

We now estimate $\boldsymbol{\lambda}$ using Simulated-GMM (SGMM). Define,

$$\hat{\boldsymbol{\lambda}}_{T,S,J,N} = \arg\min_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}} G_{T,S,J,N}(\hat{\boldsymbol{\theta}}_T(\boldsymbol{\lambda}),\boldsymbol{\lambda})^\top \mathbf{W}_{T,N}^{-1} G_{T,S,J,N}(\hat{\boldsymbol{\theta}}_T(\boldsymbol{\lambda}),\boldsymbol{\lambda}) \equiv \arg\min_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}} Z_{T,S,J,N}(\boldsymbol{\lambda}),$$

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

$$Z_{T,S,J,N}^{\dagger}(\boldsymbol{\lambda}) = G_{T,S,J,N}(\boldsymbol{\theta}^{\dagger}(\boldsymbol{\lambda}),\boldsymbol{\lambda})^\top \mathbf{W}_{T,N}^{-1} G_{T,S,J,N}(\boldsymbol{\theta}^{\dagger}(\boldsymbol{\lambda}),\boldsymbol{\lambda})$$

and

$$\boldsymbol{\lambda}^{\dagger} = \arg\min_{\boldsymbol{\lambda}} Z_{\infty}^{\dagger}(\boldsymbol{\lambda}), \text{ where } Z_{\infty}^{\dagger}(\boldsymbol{\lambda}) = \text{plim}_{T,S,J,N\to\infty} Z_{T,S,J,N}^{\dagger}(\boldsymbol{\lambda}),$$

The difficulty lies in the fact that the moment conditions depends on estimated parameters, i.e. on $\hat{\boldsymbol{\theta}}_T(\boldsymbol{\lambda})$, 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 $\sqrt{T}G_{T,S,J,N}(\hat{\boldsymbol{\theta}}_T(\boldsymbol{\lambda}^{\dagger}),\boldsymbol{\lambda}^{\dagger})$ will reflect the contribution of the asymptotic covariance of $\sqrt{T}(\hat{\boldsymbol{\theta}}_T(\boldsymbol{\lambda}) - \boldsymbol{\theta}^{\dagger}(\boldsymbol{\lambda}))$, which is not known in closed form.

Consider the following conditions.

Assumptions:
A4: $Z_{\infty}^{\dagger}(\boldsymbol{\lambda}) \le Z_{\infty}^{*}(\boldsymbol{\lambda}')$ for all $\boldsymbol{\lambda} \neq \boldsymbol{\lambda}'$.
A5: $\boldsymbol{W}_{T,N}^{-1} - \boldsymbol{W}_{\infty}^{-1} = o_P(1)$.

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,*

$$\sqrt{T}G_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}_T(\boldsymbol{\lambda}^{\dagger}),\boldsymbol{\lambda}^{\dagger}\right) \xrightarrow{d} \mathrm{N}(\mathbf{0},\boldsymbol{V}),$$

where

$$\boldsymbol{V} = \lim_{T\to\infty} \text{var}\left[\mathbf{A}_{\boldsymbol{\theta}^{\dagger}}^{\dagger}\sqrt{T}\left(\hat{\boldsymbol{\theta}}_T(\boldsymbol{\lambda}^{\dagger}) - \boldsymbol{\theta}^{\dagger}(\boldsymbol{\lambda}^{\dagger})\right)\right] + \lim_{T,S,J,N\to\infty} \text{var}\left[\sqrt{T}G_{T,S,J,N}\left(\boldsymbol{\theta}^{\dagger}(\boldsymbol{\lambda}^{\dagger}),\boldsymbol{\lambda}^{\dagger}\right)\right]$$
$$+ \lim_{T,S,J,N\to\infty} \text{cov}\left[\sqrt{T}G_{T,S,J,N}\left(\boldsymbol{\theta}^{\dagger}(\boldsymbol{\lambda}^{\dagger}),\boldsymbol{\lambda}^{\dagger}\right), \mathbf{A}_{\boldsymbol{\theta}^{\dagger}}^{\dagger}\sqrt{T}\left(\hat{\boldsymbol{\theta}}_T(\boldsymbol{\lambda}^{\dagger}) - \boldsymbol{\theta}^{\dagger}(\boldsymbol{\lambda}^{\dagger})\right)\right], \quad (26)$$

and $\boldsymbol{A}^{\dagger}_{\boldsymbol{\theta}^{\dagger}} = \text{plim}_{T,S,J,N\to\infty} \nabla_{\boldsymbol{\theta}} G_{T,S,J,N}\left(\boldsymbol{\theta}^{\dagger}\left(\boldsymbol{\lambda}^{\dagger}\right), \boldsymbol{\lambda}^{\dagger}\right)$.

It is immediate to see that the limiting distribution of $\sqrt{T} G_{T,S,J,N}(\hat{\boldsymbol{\theta}}_T\left(\boldsymbol{\lambda}^{\dagger}\right), \boldsymbol{\lambda}^{\dagger})$ reflects the contribution of the asymptotic covariance of $\sqrt{T}(\hat{\boldsymbol{\theta}}_T(\boldsymbol{\lambda}) - \boldsymbol{\theta}^{\dagger}(\boldsymbol{\lambda}))$, which is not known in closed form. For this reason, $\boldsymbol{W}_{T,N}$ cannot be an optimal weighting matrix. In fact, $\boldsymbol{W}_{\infty} = \text{plim}_{T,N\to\infty} \boldsymbol{W}_{T,N}$ is the inverse of the long-run variance of $\sqrt{T} G_{T,S,J,N}\left(\boldsymbol{\theta}^{\dagger}\left(\boldsymbol{\lambda}^{\dagger}\right), \boldsymbol{\lambda}^{\dagger}\right)$ which differs from the long-run variance of $\sqrt{T} G_{T,S,J,N}(\hat{\boldsymbol{\theta}}_T\left(\boldsymbol{\lambda}^{\dagger}\right), \boldsymbol{\lambda}^{\dagger})$. As a consequence, $\hat{\boldsymbol{\lambda}}_{T,S,J,N}$ 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,*

$$\sqrt{T}\left(\hat{\boldsymbol{\lambda}}_{T,S,J,N} - \boldsymbol{\lambda}^{\dagger}\right) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \left(\boldsymbol{A}^{\dagger\top}_{\boldsymbol{\lambda}^{\dagger}} \boldsymbol{W}^{-1}_{\infty} \boldsymbol{A}^{\dagger}_{\boldsymbol{\lambda}^{\dagger}}\right)^{-1} \boldsymbol{A}^{\dagger\top}_{\boldsymbol{\lambda}^{\dagger}} \boldsymbol{W}^{-1}_{\infty} \boldsymbol{V} \boldsymbol{W}^{-1}_{\infty} \boldsymbol{A}^{\dagger\top}_{\boldsymbol{\lambda}^{\dagger}} \left(\boldsymbol{A}^{\dagger\top}_{\boldsymbol{\lambda}^{\dagger}} \boldsymbol{W}^{-1}_{\infty} \boldsymbol{A}^{\dagger}_{\boldsymbol{\lambda}^{\dagger}}\right)^{-1}\right),$$

where $\boldsymbol{A}^{\dagger}_{\boldsymbol{\lambda}^{\dagger}} = \text{plim}_{T,S,J,N\to\infty} \nabla_{\boldsymbol{\lambda}} G_{T,S,J,N}\left(\boldsymbol{\theta}^{\dagger}\left(\boldsymbol{\lambda}^{\dagger}\right), \boldsymbol{\lambda}^{\dagger}\right)$, *and* $\boldsymbol{V}$ *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 $\boldsymbol{V}^{-1}$. The critical issue, however, is that there is no available estimator of $\boldsymbol{V}$, as

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

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 $\mathbb{E}\left[IV_{t,t+1} | \mathcal{F}_t\right]$ is also strong mixing, provided $\mathcal{F}_t$ 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.[5]

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

$$w^*_1(\boldsymbol{\theta}, \boldsymbol{\lambda}), \cdots, w^*_l(\boldsymbol{\theta}, \boldsymbol{\lambda}), w^*_{l+1}(\boldsymbol{\theta}, \boldsymbol{\lambda}), \cdots, w^*_T(\boldsymbol{\theta}, \boldsymbol{\lambda})$$

is equal to

$$w_{I_1+1}(\boldsymbol{\theta}, \boldsymbol{\lambda}), \cdots, w_{I_1+l}(\boldsymbol{\theta}, \boldsymbol{\lambda}), w_{I_2}(\boldsymbol{\theta}, \boldsymbol{\lambda}), \cdots, w_{I_b+l}(\boldsymbol{\theta}, \boldsymbol{\lambda}),$$

[5] 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 θ and λ.

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

$$\hat{\boldsymbol{\theta}}^*_T(\hat{\boldsymbol{\lambda}}_{T,S,J,N}) = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{1}{T}\sum_{t=1}^{T} l^*_t(\boldsymbol{\theta}, \hat{\boldsymbol{\lambda}}_{T,S,J,N}),$$

and

$$G^*_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}^*_T(\hat{\boldsymbol{\lambda}}_{T,S,J,N}), \hat{\boldsymbol{\lambda}}_{T,S,J,N}\right) = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{J}\sum_{j=1}^{J} g_{j,S}\left(\hat{\boldsymbol{\theta}}^*_T(\hat{\boldsymbol{\lambda}}_{T,S,J,N}), \hat{\boldsymbol{\lambda}}_{T,S,J,N}\right) - g^*_{t,N}\right).$$

Note that we have not resampled $g_{j,S}\left(\boldsymbol{\theta}\left(\boldsymbol{\lambda}\right), \boldsymbol{\lambda}\right)$; in fact for $S/T \to \infty$ and $J/T \to \infty$, the simulation error is negligible, and

$$\sqrt{T}\left(\frac{1}{J}\sum_{j=1}^{J} g_{j,S}\left(\boldsymbol{\theta}\left(\boldsymbol{\lambda}\right), \boldsymbol{\lambda}\right) - \mathrm{E}\left(g_{j,S}\left(\boldsymbol{\theta}\left(\boldsymbol{\lambda}\right), \boldsymbol{\lambda}\right)\right)\right) = o_P(1).$$

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

$$\bar{G}_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}_T(\hat{\boldsymbol{\lambda}}_{T,S,J,N}), \hat{\boldsymbol{\lambda}}_{T,S,J,N}\right) = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{J}\sum_{j=1}^{J} g_{j,S}\left(\hat{\boldsymbol{\theta}}_T(\hat{\boldsymbol{\lambda}}_{T,S,J,N}), \hat{\boldsymbol{\lambda}}_{T,S,J,N}\right) - g_{t,N}\right),$$

and note that it denotes the mean of the boostrap moment conditions, which is of order $O_P(T^{-1/2})$ in the overidentified case.[6] Therefore, in the overidentified case, recentering is necessary even for first order validity.

Basically, it can be shown that the limiting distribution of

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

is the same as the limiting distribution of $\sqrt{T}G_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}_T(\boldsymbol{\lambda}^\dagger), \boldsymbol{\lambda}^\dagger\right)$, 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

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

as $T, S, J, N \to \infty$, will approach

$$\lim_{T,S,J,N\to\infty} \mathrm{var}\left[\sqrt{T}G_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}_T(\boldsymbol{\lambda}^\dagger), \boldsymbol{\lambda}^\dagger\right)\right].$$

[6]In the exactly identified case, $\bar{G}_{T,S,J,N}(\hat{\boldsymbol{\theta}}_T(\hat{\boldsymbol{\lambda}}_{T,S,J,N}), \hat{\boldsymbol{\lambda}}_{T,S,J,N})$ is instead $O_P(l/T)$, and so recentering is not necessary for the first order validity of the bootstrap.

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,

$$\begin{aligned}
&\boldsymbol{V}_{T,B}\left(\hat{\boldsymbol{\theta}}_T(\hat{\boldsymbol{\lambda}}_{T,S,J,N}),\hat{\boldsymbol{\lambda}}_{T,S,J,N}\right)\\
&=\frac{T}{B}\sum_{i=1}^{B}\left[\left(G^*_{T,S,J,N,i}\left(\hat{\boldsymbol{\theta}}^*_{T,i}(\hat{\boldsymbol{\lambda}}_{T,S,J,N}),\hat{\boldsymbol{\lambda}}_{T,S,J,N}\right)-\bar{G}_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}_T(\hat{\boldsymbol{\lambda}}_{T,S,J,N}),\hat{\boldsymbol{\lambda}}_{T,S,J,N}\right)\right)^\top\right.\\
&\left.\left(G^*_{T,S,J,N,i}\left(\hat{\boldsymbol{\theta}}^*_{T,i}(\hat{\boldsymbol{\lambda}}_{T,S,J,N}),\hat{\boldsymbol{\lambda}}_{T,S,J,N}\right)-\bar{G}_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}_T(\hat{\boldsymbol{\lambda}}_{T,S,J,N}),\hat{\boldsymbol{\lambda}}_{T,S,J,N}\right)\right)\right],
\end{aligned}$$

and as $B\to\infty$, $\boldsymbol{V}_{T,B}\left(\hat{\boldsymbol{\theta}}_T(\hat{\boldsymbol{\lambda}}_{T,S,J,N}),\hat{\boldsymbol{\lambda}}_{T,S,J,N}\right)$ approaches

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

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 $B\to\infty$,*

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

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

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

3.5 An efficient Concentrated SGMM estimator

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

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

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 $B\to\infty$,*

$$\sqrt{T}\left(\tilde{\boldsymbol{\lambda}}_{T,S,J,N}-\boldsymbol{\lambda}^\dagger\right)\xrightarrow{d}\mathrm{N}\left(\mathbf{0},\left(\boldsymbol{A}^{\dagger\top}_{\boldsymbol{\lambda}^\dagger}\boldsymbol{V}\boldsymbol{A}^\dagger_{\boldsymbol{\lambda}^\dagger}\right)^{-1}\right),$$

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

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

$$
\begin{aligned}
&\tilde{\boldsymbol{\lambda}}^{*}_{T,S,J,N} \\
&= \arg\min_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}} \left[\left(G^{*}_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}^{*}_{T}(\boldsymbol{\lambda}),\boldsymbol{\lambda}\right) - \bar{G}_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}_{T}(\tilde{\boldsymbol{\lambda}}_{T,S,J,N}),\tilde{\boldsymbol{\lambda}}_{T,S,J,N}\right)\right)^{\top} \right. \\
&\quad \left. \boldsymbol{V}_{T,B}\left(\hat{\boldsymbol{\theta}}_{T}(\hat{\boldsymbol{\lambda}}_{T,S,J,N}),\hat{\boldsymbol{\lambda}}_{T,S,J,N}\right)^{-1}\left(G^{*}_{T,S,J,N}(\hat{\boldsymbol{\theta}}^{*}_{T}(\lambda),\boldsymbol{\lambda}) - \bar{G}_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}_{T}(\tilde{\boldsymbol{\lambda}}_{T,S,J,N}),\tilde{\boldsymbol{\lambda}}_{T,S,J,N}\right)\right)\right].
\end{aligned}
$$

We have,

Theorem 6 *Let A1-A5 hold. If $S/J\to\infty$, $\xi^{-1}/S\to\infty$, $J/T\to\infty$, $N/T\to\infty, l/T^{1/2}\to 0$ and $l/T^{1/2}\to 0$, then as $T\to\infty$ and $B\to\infty$,*

$$
P\left(\omega : \sup_{x\in\mathbb{R}}\left|P^{*}\left[\sqrt{T}\left(\tilde{\boldsymbol{\lambda}}^{*}_{T,S,J,N}-\tilde{\boldsymbol{\lambda}}_{T,S,J,N}\right)\le x\right] - P\left[\sqrt{T}\left(\left(\tilde{\boldsymbol{\lambda}}_{T,S,J,N}-\boldsymbol{\lambda}^{\dagger}\right)\right)\le x\right]\right| > \varepsilon\right)\to 0,
$$

where P^{*} is the probability law governing the bootstrap.

4 Empirical analysis

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 $T = 672$ observations. We take these two series to compute the two macroeconomic factors, inflation and the growth rate for the industrial production, both at a yearly level,

$$
y_{1,t} = \text{Inflation}_{t-12\to t} = \text{CPI}_t/\text{CPI}_{t-12} \quad \text{and} \quad y_{2,t} = \text{Growth}_{t-12\to t} = \text{IP}_t/\text{IP}_{t-12},
$$

where CPI_t is the consumer price index and IP_t is the industrial production index, as of month t. Figure 1 depicts the two series $y_{1,t}$ and $y_{2,t}$, along with the NBER recession events.

To form moment conditions, we compute the log-price variations as $R_t = \log(s_t/s_{t-12})$ and, following Mele (2007) and Fornari and Mele (2007), price volatility as,

$$
\text{Vol}_t = \sqrt{6\pi}\cdot\frac{1}{12}\sum_{i=1}^{12}\left|R^m_{t+1-i}\right|,
$$

where R^m_t is the monthly log-price variation, defined as $R^m_t = \log(s_t/s_{t-1})$.

Consider, then, the following regressions,

$$
R_t = a^R + b^R_{1,12}y_{1,t-12} + b^R_{2,12}y_{2,t-12} + \epsilon^R_t,
$$

and,

$$\mathrm{Vol}_t = a^V + \sum_{i\in\{6,12,18,24,36,48\}} \phi_i \mathrm{Vol}_{t-i} + \sum_{i\in\{12,24,36,48\}} b^V_{1,i} y_{1,t-i} + \sum_{i\in\{12,24,36,48\}} b^V_{2,i} y_{2,t-i} + \epsilon^V_t,$$

where ϵ^R_t and ϵ^V_t are residual terms. Define μ_S to be the mean of the real stock price, s, and V_S to be the mean of return volatility, Vol. The parameter vector we want to match is

$$\boldsymbol{\psi} = \left(a^R, b^R_{1,12}, b^R_{2,12}, a^V, (\phi_i)_{i\in\{6,12,18,24,36,48\}}, \left(b^V_{1,i}\right)_{i\in\{12,24,36,48\}}, \left(b^V_{2,i}\right)_{i\in\{12,24,36,48\}}, \mu_S, V_S\right).$$

Finally, we fix the parameter s_0 to,

$$s_0 = \mu_S - s_1\bar{y}_1 - s_2\bar{y}_2 - s_3\mu_3.$$

Table 1 reports the moment conditions. Tables 2 and 3 report parameter estimates. Figure 2 depicts the dynamics of the log-price changes and the volatility of those.

5 Conclusion

Appendix

A. Proofs for Section 2

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

(i) For all $\boldsymbol{y} : \boldsymbol{V}(\boldsymbol{y})_{(ii)} = 0$, $\boldsymbol{\beta}_i^\top (-\boldsymbol{\kappa}\boldsymbol{y} + \boldsymbol{\kappa}\boldsymbol{\mu}) > \frac{1}{2}\boldsymbol{\beta}_i^\top \boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top \boldsymbol{\beta}_i$

(ii) For all j, if $\left(\boldsymbol{\beta}_i^\top \boldsymbol{\Sigma}\right)_j \neq 0$, then $\boldsymbol{V}_{ii} = \boldsymbol{V}_{jj}$.

Then, by Duffie and Kan (1996) (unnumbered theorem, p. 388), there is a unique strong solution to (1) for which $\boldsymbol{V}(\boldsymbol{y}(t))_{(ii)} > 0$ for all t almost surely.

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

$$\text{For all } y_i : \alpha_i + \beta_i y_i = 0, \quad \begin{cases} -\beta_1 \left[\kappa_1 (y_1 - \mu_1) + \bar{\kappa}_1 (y_2 - \mu_2)\right] > \dfrac{1}{2}\beta_1^2 \\ -\beta_2 \left[\kappa_2 (y_2 - \mu_2) + \bar{\kappa}_2 (y_1 - \mu_1)\right] > \dfrac{1}{2}\beta_2^2 \\ -\beta_3 (\kappa_3 y_3 - \kappa_3 \mu_3) > \dfrac{1}{2}\beta_3^2 \end{cases}$$

This is, ruling out the trivial case in which $\beta_i = 0$,

$$\begin{cases} \kappa_1 (\alpha_1 + \mu_1 \beta_1) + \bar{\kappa}_1 \beta_1 \left(\dfrac{\alpha_2}{\beta_2} + \mu_2\right) > \dfrac{1}{2}\beta_1^2 \\ \kappa_2 (\alpha_2 + \mu_2 \beta_2) + \bar{\kappa}_2 \beta_2 \left(\dfrac{\alpha_1}{\beta_1} + \mu_1\right) > \dfrac{1}{2}\beta_2^2 \\ \kappa_3 (\alpha_3 + \mu_3 \beta_3) > \dfrac{1}{2}\beta_3^2 \end{cases}$$

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

$$\frac{\mathrm{d}s^\xi(t)}{s^\xi(t)} = \left[-r + \frac{\mathcal{L}s_y(\boldsymbol{y}(t))}{s(t)} - \frac{\boldsymbol{s}_y(\boldsymbol{y}(t))^\top \boldsymbol{\Sigma}\boldsymbol{V}(\boldsymbol{y}(t))\boldsymbol{\Lambda}(t)}{s(t)}\right]\mathrm{d}t + \left[\frac{s_y(\boldsymbol{y}(t))^\top \boldsymbol{\Sigma}\boldsymbol{V}(\boldsymbol{y}(t))}{s(t)} - \boldsymbol{\Lambda}(t)^\top\right]\mathrm{d}\boldsymbol{W}(t) \tag{A1}$$

where

$$\mathcal{L}s(\boldsymbol{y}) \equiv s_y(\boldsymbol{y})^\top \boldsymbol{\kappa}(\boldsymbol{\mu} - \boldsymbol{y}) + \frac{1}{2}\mathrm{tr}\left\{[\boldsymbol{\Sigma}\boldsymbol{V}(\boldsymbol{y})][\boldsymbol{\Sigma}\boldsymbol{V}(\boldsymbol{y})]^\top s_{yy}(\boldsymbol{y})\right\}.$$

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

$$s^{\xi}(t) = E\left[\int_t^{\tau} \delta^{\xi}(h)\, dh \,\middle|\, F(t)\right] + E[s^{\xi}(\tau) \mid F(t)], \tag{A2}$$

where $\delta^{\xi}(\tau)$ 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,

$$\lim_{T\to\infty} E[s^{\xi}(T) \mid F(t)] = 0, \tag{A3}$$

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

Next, by Eq. (A2),

$$0 = \frac{\mathrm{d}}{\mathrm{d}\tau} E[s^{\xi}(\tau) \mid F(t)]\Big|_{\tau=t} + \delta^{\xi}(t). \tag{A4}$$

Below, we show that

$$E[s^{\xi}(\tau) \mid F(t)] = s^{\xi}(t) + \int_t^{\tau} \left[-r + \frac{\mathcal{L}s(\boldsymbol{y}(h))}{s(h)} - \frac{s_y(\boldsymbol{y}(h))^{\top}\boldsymbol{\Sigma}\boldsymbol{V}(\boldsymbol{y}(h))\boldsymbol{\Lambda}(h)}{s(h)}\right] s^{\xi}(h)\,\mathrm{d}h. \tag{A5}$$

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

$$\text{For all } \boldsymbol{y}, \quad s_y(\boldsymbol{y})^{\top}(\boldsymbol{c} - \boldsymbol{D}\boldsymbol{y}) + \frac{1}{2}\mathrm{tr}\left\{[\boldsymbol{\Sigma}\boldsymbol{V}(\boldsymbol{y})][\boldsymbol{\Sigma}\boldsymbol{V}(\boldsymbol{y})]^{\top} s_{yy}(\boldsymbol{y})\right\} + \delta(\boldsymbol{y}) - rs(\boldsymbol{y}) = 0, \tag{A6}$$

where $\boldsymbol{c}$ and $\boldsymbol{D}$ are defined in the proposition.

Let us assume that the price function is affine in $\boldsymbol{y}$,

$$s(\boldsymbol{y}) = \gamma + \boldsymbol{\eta}^{\top}\boldsymbol{y}, \tag{A7}$$

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

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

That is,

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

The solution to this system is,

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

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

As regards Eq. (A3), we have

$$\begin{aligned}
\lim_{T\to\infty} E[s^{\xi}(T) \mid F(t)] &= \lim_{T\to\infty} E[e^{-r(T-t)}\xi(T)\, s(\boldsymbol{y}(T)) \mid F(t)] \\
&= \gamma e^{-r(T-t)} \lim_{T\to\infty} E[\xi(T) \mid F(t)] + \lim_{T\to\infty} e^{-r(T-t)} E[\xi(T)\, \boldsymbol{\eta}^{\top}\boldsymbol{y}(T) \mid F(t)] \\
&= \xi(t) \lim_{T\to\infty} e^{-r(T-t)} \mathbb{E}[\boldsymbol{\eta}^{\top}\boldsymbol{y}(T) \mid F(t)],
\end{aligned}$$

where the second line follows by Eq. (A7), and the third line holds because $E[\xi(T) \mid F(t)] = 1$, and by a change of measure (we need some more rigorous work on this). Define $Q(t) = \boldsymbol{\eta}^{\top}\boldsymbol{y}(t)$. Under the risk-neutral probability

$$dQ(t) = \boldsymbol{\eta}^{\top}(\boldsymbol{c} - \boldsymbol{D}\boldsymbol{y}(t))\,\mathrm{d}t + \boldsymbol{\eta}^{\top}\boldsymbol{\Sigma}\boldsymbol{V}(\boldsymbol{y}(t))\,\mathrm{d}\hat{\boldsymbol{W}}(t)$$

In progress.

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

$$E\left[\int_t^T \left\| \frac{s_y(\boldsymbol{y}(\tau))^{\top}\boldsymbol{\Sigma}\boldsymbol{V}(\boldsymbol{y}(\tau))}{s(\tau)} - \boldsymbol{\Lambda}(\tau)^{\top} \right\|^2 d\tau \right] < \infty.$$

This is:

$$E\left[\int_t^T \left\| \frac{\boldsymbol{\eta}^{\top}\boldsymbol{\Sigma}\boldsymbol{V}(\boldsymbol{y}(\tau))}{\gamma + \boldsymbol{\eta}^{\top}\boldsymbol{y}(\tau)} - \boldsymbol{\Lambda}(\tau)^{\top} \right\|^2 d\tau \right] < \infty.$$

In progress. ■

B. Proofs for Section 3

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

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

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

(ii) Via a mean value expansion around $\boldsymbol{\theta}^*(\boldsymbol{\lambda})$,

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

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

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_T(\boldsymbol{\lambda}) - \boldsymbol{\theta}^{\dagger}(\boldsymbol{\lambda})\right) = \boldsymbol{B}(\boldsymbol{\theta}^{\dagger}(\boldsymbol{\lambda}))^{-1}\nabla_{\boldsymbol{\theta}} l_T(\boldsymbol{\theta}^{\dagger}(\boldsymbol{\lambda})) + o_P(1)$$

where the $o_P(1)$ holds uniformly in $\boldsymbol{\lambda}$, given that $\sup_{\boldsymbol{\theta}\times\boldsymbol{\lambda}\in\Theta\times\boldsymbol{\Lambda}}\left(\nabla^2_{\boldsymbol{\theta}} l_T\left(\boldsymbol{\theta}\left(\boldsymbol{\lambda}\right)\right)-\boldsymbol{B}\left(\boldsymbol{\theta}^*\left(\boldsymbol{\lambda}\right)\right)\right) = o_P(1)$. 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:

$$\begin{aligned}\sqrt{T}G_{T,S,J,N}(\hat{\boldsymbol{\theta}}_T(\boldsymbol{\lambda}^\dagger),\boldsymbol{\lambda}^\dagger) &= \sqrt{T}G_{T,S,J,N}(\boldsymbol{\theta}^\dagger(\boldsymbol{\lambda}^\dagger),\boldsymbol{\lambda}^\dagger)\\ &\quad+\nabla_{\boldsymbol{\theta}}G_{T,S,J,N}(\bar{\boldsymbol{\theta}}_T(\boldsymbol{\lambda}^\dagger),\boldsymbol{\lambda}^\dagger)\sqrt{T}\left(\hat{\boldsymbol{\theta}}_T(\boldsymbol{\lambda}^\dagger)-\boldsymbol{\theta}^\dagger(\boldsymbol{\lambda}^\dagger)\right)\\ &= \sqrt{T}G_{T,S,J,N}(\boldsymbol{\theta}^\dagger(\boldsymbol{\lambda}^\dagger),\boldsymbol{\lambda}^\dagger)+\boldsymbol{A}^\dagger_{\boldsymbol{\theta}^\dagger}\sqrt{T}\left(\hat{\boldsymbol{\theta}}_T(\boldsymbol{\lambda}^\dagger)-\boldsymbol{\theta}^\dagger(\boldsymbol{\lambda}^\dagger)\right)+o_P(1),\end{aligned}$$

given that, by Lemma A1,

$$\nabla_{\boldsymbol{\theta}}G_{T,S,J,N}\left(\bar{\boldsymbol{\theta}}_T(\boldsymbol{\lambda}^\dagger),\boldsymbol{\lambda}^\dagger\right)=\nabla_{\boldsymbol{\theta}}G_{T,S,J,N}\left(\boldsymbol{\theta}^\dagger(\boldsymbol{\lambda}^\dagger),\boldsymbol{\lambda}^\dagger\right)+o_P(1)$$

and $\mathrm{plim}_{T,S,J,N\to\infty}\nabla_{\boldsymbol{\theta}}G_{T,S,J}\left(\boldsymbol{\theta}^\dagger\left(\boldsymbol{\lambda}^\dagger\right),\boldsymbol{\lambda}^\dagger\right)=\boldsymbol{A}^\dagger_{\boldsymbol{\theta}^\dagger}$. The result then follows from central limit theorem.

Proof of Theorem 3: Via a mean value expansion around $\boldsymbol{\lambda}^*$,

$$\begin{aligned}0&=\sqrt{T}\nabla_{\boldsymbol{\lambda}}\hat{Z}_{T,S,J,N}(\hat{\boldsymbol{\lambda}}_{T,S,J,N})\\ &=\nabla_{\boldsymbol{\lambda}}G_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}_T(\hat{\boldsymbol{\lambda}}_{T,S,J,N}),\hat{\boldsymbol{\lambda}}_{T,S,J,N}\right)^\top\boldsymbol{W}^{-1}_{T,N}\sqrt{T}G_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}_T(\boldsymbol{\lambda}^\dagger),\boldsymbol{\lambda}^\dagger\right)\\ &+\nabla_{\boldsymbol{\lambda}}G_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}_T(\hat{\boldsymbol{\lambda}}_{T,S,J,N}),\hat{\boldsymbol{\lambda}}_{T,S,J,N}\right)^\top\boldsymbol{W}^{-1}_{T,N}\nabla_{\boldsymbol{\lambda}}G_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}_T\left(\bar{\boldsymbol{\lambda}}_{T,S,J,N}\right),\bar{\boldsymbol{\lambda}}_{T,S,J,N}\right)\\ &\times\sqrt{T}\left(\hat{\boldsymbol{\lambda}}_{T,S,J,N}-\boldsymbol{\lambda}^\dagger\right).\end{aligned}$$

Thus,

$$\begin{aligned}&\sqrt{T}\left(\hat{\boldsymbol{\lambda}}_{T,S,J,N}-\boldsymbol{\lambda}^\dagger\right)\\ &=-\left(\nabla_{\boldsymbol{\lambda}}G_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}_T(\hat{\boldsymbol{\lambda}}_{T,S,J,N}),\hat{\boldsymbol{\lambda}}_{T,S,J,N}\right)^\top\boldsymbol{W}^{-1}_{T,N}\nabla_{\boldsymbol{\lambda}}G_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}_T(\bar{\boldsymbol{\lambda}}_{T,S,J,N}),\bar{\boldsymbol{\lambda}}_{T,S,J,N}\right)\right)^{-1}\\ &\times\nabla_{\boldsymbol{\lambda}}G_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}_T(\hat{\boldsymbol{\lambda}}_{T,S,J,N}),\hat{\boldsymbol{\lambda}}_{T,S,J,N}\right)^\top\boldsymbol{W}^{-1}_{T,N}G_{T,S,J}\left(\hat{\boldsymbol{\theta}}_T(\boldsymbol{\lambda}^\dagger),\boldsymbol{\lambda}^\dagger\right)\end{aligned}$$

Given Theorem 1 and Lemma A1,

$$\mathrm{plim}_{T,S,J\to\infty}\nabla_{\boldsymbol{\lambda}}G_{T,S,J}\left(\hat{\boldsymbol{\theta}}_T(\hat{\boldsymbol{\lambda}}_{T,S,J}),\hat{\boldsymbol{\lambda}}_{T,S,J,N}\right)^\top=\boldsymbol{A}^\dagger_{\boldsymbol{\lambda}^\dagger}.$$

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,

$$P\left(\omega:\sup_{x\in\Re}\left|P^*\left[\sqrt{T}\left(G^*_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}^*_T(\hat{\boldsymbol{\lambda}}_{T,S,J,N}),\hat{\boldsymbol{\lambda}}_{T,S,J,N}\right)-\bar{G}_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}_T(\hat{\boldsymbol{\lambda}}_{T,S,J,N}),\hat{\boldsymbol{\lambda}}_{T,S,J,N}\right)\right)\leq x\right]\right.\right.$$
$$\left.\left.-\ P\left[\sup_{u\times v\in U\times V}\sqrt{T}G_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}_T(\boldsymbol{\lambda}^\dagger),\boldsymbol{\lambda}^\dagger\right)\leq x\right]\right|>\varepsilon\right)\to 0$$

Step 2: We show that,

$$\mathrm{E}^*\left[\left|\sqrt{T}\left(G^*_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}^*_T(\hat{\boldsymbol{\lambda}}_{T,S,J,N}),\hat{\boldsymbol{\lambda}}_{T,S,J,N}\right)-\bar{G}_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}_T(\hat{\boldsymbol{\lambda}}_{T,S,J,N}),\hat{\boldsymbol{\lambda}}_{T,S,J,N}\right)\right)\right|^{2+\delta}\right]<\infty,$$

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 $T\to\infty$,

$$\mathrm{var}^*\left[\sqrt{T}\left(G^*_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}^*_T(\hat{\boldsymbol{\lambda}}_{T,S,J,N}),\hat{\boldsymbol{\lambda}}_{T,S,J,N}\right)-\bar{G}_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}_T(\hat{\boldsymbol{\lambda}}_{T,S,J,N}),\hat{\boldsymbol{\lambda}}_{T,S,J,N}\right)\right)\right]$$
$$\overset{pr}{\to}\boldsymbol{V},$$

where var^* denotes the variance according to the bootstrap probability law, conditional on the sample. As $B\to\infty$,

$$\boldsymbol{V}_{T,B}\left(\hat{\boldsymbol{\theta}}_T(\hat{\boldsymbol{\lambda}}_{T,S,J,N}),\hat{\boldsymbol{\lambda}}_{T,S,J,N}\right)$$
$$\overset{pr^*}{\to}\mathrm{var}^*\left[\sqrt{T}\left(G^*_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}^*_T(\hat{\boldsymbol{\lambda}}_{T,S,J,N}),\hat{\boldsymbol{\lambda}}_{T,S,J,N}\right)-\bar{G}_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}_T(\hat{\boldsymbol{\lambda}}_{T,S,J,N}),\hat{\boldsymbol{\lambda}}_{T,S,J,N}\right)\right)\right],$$

the statement above follows.

Proof of Step 1:

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

We begin by showing that

$$P\left(\omega:\sup_{x\in\Re}P^*\left[\sqrt{T}\left(\hat{\boldsymbol{\theta}}^*_T(\hat{\boldsymbol{\lambda}}_{T,S,J,N})-\hat{\boldsymbol{\theta}}_T(\hat{\boldsymbol{\lambda}}_{T,S,J,N})\right)\leq x\right]-P\left[\sqrt{T}\left(\hat{\boldsymbol{\theta}}_T(\boldsymbol{\lambda}^\dagger)-\boldsymbol{\theta}^\dagger(\boldsymbol{\lambda}^\dagger)\right)\leq x\right]>\varepsilon\right)$$
$$\to 0.$$

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

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

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

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

Now,

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

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

$$\begin{aligned}
&\mathrm{var}^*\left[\sqrt{T}\nabla_{\boldsymbol{\theta}} l_T^*(\hat{\boldsymbol{\theta}}_T, \hat{\boldsymbol{\lambda}}_{T,S,J,N})\right] \\
&= \mathrm{var}^*\left[\frac{1}{\sqrt{T}}\sum_{k=1}^{b}\sum_{i=1}^{l_T}\nabla_{\boldsymbol{\theta}} l_{I_K+i}^*(\hat{\boldsymbol{\theta}}_T, \hat{\boldsymbol{\lambda}}_{T,S,J,N})\right] \\
&= \mathrm{E}^*\left[\frac{1}{T}\sum_{k=1}^{b}\sum_{i=1}^{l_T}\sum_{j=1}^{l_T}\nabla_{\boldsymbol{\theta}} l_{I_K+i}(\hat{\boldsymbol{\theta}}_T, \hat{\boldsymbol{\lambda}}_{T,S,J,N})\nabla_{\boldsymbol{\theta}} l_{I_K+j}(\hat{\boldsymbol{\theta}}_T, \hat{\boldsymbol{\lambda}}_{T,S,J,N})^\top\right] \\
&= \frac{1}{T-l_T+1}\frac{1}{l}\sum_{t=0}^{T-l_T}\sum_{i=1}^{l_T}\sum_{j=1}^{l_T}\nabla_{\boldsymbol{\theta}} l_{t+i}(\hat{\boldsymbol{\theta}}_T, \hat{\boldsymbol{\lambda}}_{T,S,J,N})\nabla_{\boldsymbol{\theta}} l_{t+j}(\hat{\boldsymbol{\theta}}_T, \hat{\boldsymbol{\lambda}}_{T,S,J,N})^\top \quad \mathrm{Pr}\text{-}P \\
&= \frac{1}{T}\sum_{t=l}^{T-l}\sum_{i=-l}^{l}\nabla_{\boldsymbol{\theta}} l_t(\hat{\boldsymbol{\theta}}_T, \hat{\boldsymbol{\lambda}}_{T,S,J,N})\nabla_{\boldsymbol{\theta}} l_{t+l}^*(\hat{\boldsymbol{\theta}}_T, \hat{\boldsymbol{\lambda}}_{T,S,J,N})^\top \quad \mathrm{Pr}\text{-}P,
\end{aligned}$$

by Theorem 3.5 in Kunsch (1989). Finally,

$$\left[-\nabla_{\boldsymbol{\theta}}^2 l_T^*(\bar{\boldsymbol{\theta}}_T^*, \hat{\boldsymbol{\lambda}}_{T,S,J,N})\right]^{-1} - \mathrm{E}\left[-\nabla_{\boldsymbol{\theta}}^2 l_t(\boldsymbol{\theta}^\dagger, \boldsymbol{\lambda}^\dagger)\right] = o_P^*(1) \quad \mathrm{Pr}\text{-}P,$$

where with the notation $o_P^*(1)$ 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.

TO BE COMPLETED

Proof of Lemma A3: Straigthforward from Lemma A2 and Theorem 3.

Proof of Theorem 3: For notation brevity, let

$$G_{T,S,J,N}^{*,c}\left(\hat{\boldsymbol{\theta}}_T^*(\boldsymbol{\lambda}), \boldsymbol{\lambda}\right) = G_{T,S,J,N}^*\left(\hat{\boldsymbol{\theta}}_T^*(\boldsymbol{\lambda}), \boldsymbol{\lambda}\right) - \bar{G}_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}_T(\boldsymbol{\lambda}), \boldsymbol{\lambda}\right)$$

Via a mean value expansion around $\tilde{\boldsymbol{\lambda}}_{T,S,J,N}$,

$$
\begin{aligned}
&\sqrt{T}\left(\tilde{\boldsymbol{\lambda}}^*_{T,S,J,N} - \tilde{\boldsymbol{\lambda}}_{T,S,J,N}\right) \\
=&\Bigg[\nabla_\lambda G^{*,c}_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}^*_T(\tilde{\boldsymbol{\lambda}}^*_{T,S,J,N}), \tilde{\boldsymbol{\lambda}}^*_{T,S,J,N}\right) \boldsymbol{V}_{T,B}\left(\hat{\boldsymbol{\theta}}_T(\hat{\boldsymbol{\lambda}}_{T,S,J,N}), \hat{\boldsymbol{\lambda}}_{T,S,J,N}\right)^{-1} \\
&\times \nabla_\lambda G^{*,c}_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}^*_T(\bar{\boldsymbol{\lambda}}^*_{T,S,J,N}), \bar{\boldsymbol{\lambda}}^*_{T,S,J,N}\right)\Bigg] \\
&\times \sqrt{T}\Bigg[\nabla_\lambda G^{*,c}_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}^*_T(\tilde{\boldsymbol{\lambda}}^*_{T,S,J,N}), \tilde{\boldsymbol{\lambda}}^*_{T,S,J,N}\right) \boldsymbol{V}_{T,B}\left(\hat{\boldsymbol{\theta}}_T(\hat{\boldsymbol{\lambda}}_{T,S,J,N}), \hat{\boldsymbol{\lambda}}_{T,S,J,N}\right)^{-1} \\
&\times G^{*,c}_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}^*_T(\tilde{\boldsymbol{\lambda}}_{T,S,J,N}), \tilde{\boldsymbol{\lambda}}_{T,S,J,N}\right)\Bigg].
\end{aligned}
$$

Now, by the same argument used in the proof of Theorem 3, $\sqrt{T} G^{*,c}_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}^*_T(\tilde{\boldsymbol{\lambda}}_{T,S,J,N}), \tilde{\boldsymbol{\lambda}}_{T,S,J,N}\right)$ has the same limiting distribution as $\sqrt{T} G_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}_T(\boldsymbol{\lambda}^\dagger), \boldsymbol{\lambda}^\dagger\right)$. Furthermore,

$$
\nabla_\lambda G^{*,c}_{T,S,J,N}\left(\hat{\boldsymbol{\theta}}^*_T(\tilde{\boldsymbol{\lambda}}^*_{T,S,J,N}), \tilde{\boldsymbol{\lambda}}^*_{T,S,J,N}\right) - \boldsymbol{A}^\dagger_{\boldsymbol{\lambda}^\dagger} = o^*_P(1) \quad \mathrm{Pr}\text{-}P,
$$

where

$$
\boldsymbol{A}^\dagger_{\boldsymbol{\lambda}^\dagger} = \mathrm{plim}_{T,S,J\to\infty} \nabla_{\boldsymbol{\lambda}} G_{T,S,J,N}\left(\boldsymbol{\theta}^\dagger(\boldsymbol{\lambda}^\dagger), \boldsymbol{\lambda}^\dagger\right).
$$

The statement then follows from Theorem 3.

References

Ang, A. and J. Liu, 2004, "How to Discount Cashflows with Time-Varying Expected Returns," *Journal of Finance* 59, 2745-2783.

Ang, A. and M. Piazzesi, 2003, "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables," *Journal of Monetary Economics*, 50, 745-787.

Ang, A., M. Piazzesi and M. Wei, 2005, "What Does the Yield Curve Tell Us about GDP Growth?" *Journal of Econometrics*, forthcoming.

Bansal, R. and A. Yaron, 2004, "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance* 59, 1481-1509.

Bakshi, G. and D. Madan, 2000, "Spanning and Derivative Security Evaluation," *Journal of Financial Economics* 55, 205-238.

Bakshi, G. and D. Madan, 2006, "A Theory of Volatility Spreads," forthcoming in *Management Science.*

Bekaert, G. and S.R. Grenadier, 2001, "Stock and Bond Pricing in an Affine Economy," working paper Columbia Business School.

Bernanke, B.S., M. Gertler and S. Gilchrist, 1999, "The Financial Accelerator in a Quantitative Business Cycle Framework," in J.B. Taylor and M. Woodford (Eds.): *Handbook of Macroeconomics*, Vol. 1C, Chapter 21, p. 1341-1393.

Bollerslev, T. and H. Zhou, 2005, Volatility Puzzles: A Simple Framework for Gauging Return-Volatility Regressions, *Journal of Econometrics,* forthcoming.

Bollerslev, T., M. Gibson and H. Zhou, 2004, "Dynamic Estimation of Volatility Risk Premia and Investor Risk Aversion from Option Implied and Realized Volatilities," Federal Reserve Board WP 2004-56.

Brandt, M.W., Kang, Q., 2004, "On the Relationship Between the Conditional Mean and Volatility of Stock Returns: a Latent VAR Approach," *Journal of Financial Economics* 72, 217-257.

Britten-Jones, M. and A. Neuberger, 2000, "Option Prices, Implied Price Processes and Stochastic Volatility," *Journal of Finance* 55, 839-866.

Campbell, J. Y., and J. H. Cochrane, 1999, "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy* 107, 205-251.

Campbell, J. Y., 2003, "Consumption-Based Asset Pricing," in Constantinides, G.M., M. Harris and R. M. Stulz (Editors): *Handbook of the Economics of Finance* (Volume 1B, chapter 13), 803-887.

Carr, P. and D. Madan, 2001, "Optimal Positioning in Derivative Securities," *Quantitative Finance* 1, 19-37.

Carr, P. and L. Wu, 2004, "Variance Risk Premia," Working Paper, Bloomberg L.P. and Courant Institute.

Chernov, M. and E. Ghysels, 2000, "A Study towards a Unified Approach to the Joint Estimation of Objective and Risk-Neutral Measures for the Purpose of Options Valuation," *Journal of Financial Economics* 56, 407-458.

Cox, J.C., J.E. Ingersoll and S.A. Ross, 1985, "A Theory of the Term Structure of Interest Rates," *Econometrica* 53, 385-407.

Duffee, G.R., 2002, "Term Premia and Interest Rate Forecasts in Affine Models," *Journal of Finance* 57, 405-443.

Duffie, D. and R. Kan, 1996, "A Yield-Factor Model of Interest Rates," *Mathematical Finance* 6, 379-406.

Duffie, D., Pan, J. and K. Singleton, 2000, "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica* 68, 1343-1376.

Fornari, F. and A. Mele, 2007, "Financial Volatility and Real Economic Activity," working paper ECB and LSE.

Jiang, G. and Y. Tian, 2005, "The Model-Free Implied Volatility and Its Information Content," *Review of Financial Studies* 18, 1305-1342.

Heston, S. L., 1993, "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies* 6, 327-343.

Mamaysky, H., 2002, "A Model for Pricing Stocks and Bonds," working paper Yale School of Management.

Mehra, R. and E. C. Prescott, 2003, "The Equity Premium in Retrospect," in Constantinides, G.M., M. Harris and R. M. Stulz (Editors): *Handbook of the Economics of Finance* (Volume 1B, chapter 14), 889-938.

Mele, A., 2007, "Asymmetric Stock Market Volatility and the Cyclical Behavior of Expected Returns," *Journal of Financial Economics* 86, 446-478.

Schwert, G.W., 1989a, "Why Does Stock Market Volatility Change over Time?" *Journal of Finance* 44, 1115-1153.

Schwert, G.W., 1989b, "Business Cycles, Financial Crises and Stock Volatility," *Carnegie-Rochester Conference Series on Public Policy* 31, 83-125.

Singleton, K.J., 2001, "Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function," *Journal of Econometrics* 102, 111-141.

Tauchen, G., 2005, "Stochastic Volatility in General Equilibrium," working paper, Duke University.

Tables

Table 1 – Moment conditions

	Data	Model
a^R	2.3977	2.4373
$b^R_{1,12}$	−1.4412	−1.0090
$b^R_{2,12}$	−0.8335	−1.3512
a^V	−0.3982	0.0103
ϕ_6	1.0136	0.8732
ϕ_{12}	−0.6863	−0.5546
ϕ_{18}	0.5289	0.4313
ϕ_{24}	−0.3225	−0.1561
ϕ_{36}	0.0572	0.0327
ϕ_{48}	−0.0195	0.0169
$b^V_{1,12}$	0.0709	0.1574
$b^V_{1,24}$	−0.1090	−0.0017
$b^V_{1,36}$	0.1769	−0.0292
$b^V_{1,48}$	0.0064	−0.0642
$b^V_{2,12}$	0.0458	−0.0352
$b^V_{2,24}$	0.0967	0.0206
$b^V_{2,36}$	0.1082	−0.0156
$b^V_{2,48}$	0.0356	0.0078
μ_S	2.8842	2.9002
V_S	0.1136	0.1475

Table 2 – Parameter estimates of the bivariate diffusion

	Estimate
κ_1	0.0255
μ_1	1.0379
α_1	0.0059
β_1	-0.0054
κ_2	0.5628
μ_2	1.0388
α_2	0.0517
β_2	-0.0479
$\bar{\kappa}_1$	-0.2532
$\bar{\kappa}_2$	1.1701

Table 3 – Parameter estimates of the stock price process and the unobservable factor

	Estimate
s_0	0.2356
s_1	0.0629
s_2	2.4437
s_3	0.2196
α_3	2.3022
β_3	0.2055
μ_3	0.0493
κ_3	$4.0324 \cdot 10^{-4}$

Figures

Figure 1 – Industrial production growth $(t, t+12)$ and inflation $(t, t+12)$, with NBER dated recession periods

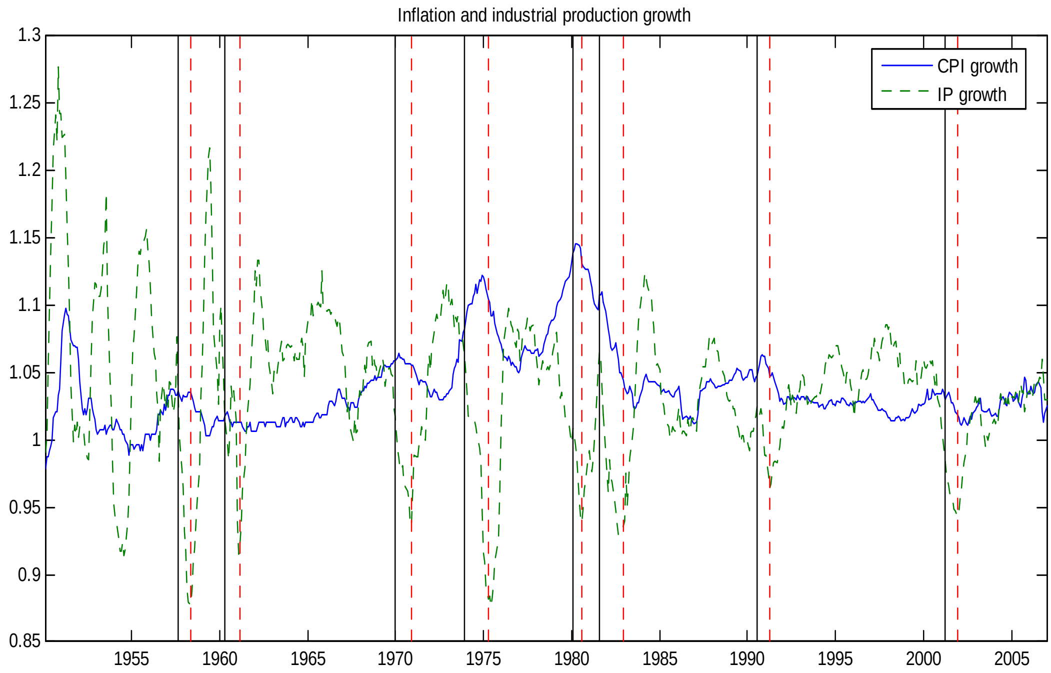

Figure 2 – Ex-post returns, ex-post volatility and model predictions