Review of Finance, 2025, **29**, 315–347
https://doi.org/10.1093/rof/rfae043
Advance Access Publication Date: 25 November 2024
Article

OXFORD

What drives commodity price variation?

Meng Han[1], Lammertjan Dam[2], Walter Pohl[3,*]

[1]Bay Area International Business School, Beijing Normal University, Zhuhai, China
[2]Faculty of Economics and Business, University of Groningen, Groningen, The Netherlands
[3]NHH Norwegian School of Economics, Bergen, Norway

*Corresponding author: NHH Norwegian School of Economics, Helleveien 30, Bergen, 5045, Norway.
Email: Walter.Pohl@nhh.no

Abstract

We investigate the importance of time-varying discount rates for commodity prices using an index based on twenty-three commodities for the period 1959–2024. We show that in commodities markets, unlike other financial markets, time variation in discount rates plays a much smaller role. Instead, prices forecast cash flows as well as discount rates. A high price for a commodity today, measured as a low percentage net convenience yield, forecasts both a high future convenience yield and a low expected return. For longer horizons, variation in percentage net convenience yields seems mainly driven by net convenience yield growth, making commodities much closer to the classical textbook view of price changes representing news about cash flows.

Keywords: convenience yield; commodity price variation; commodity return predictability.
JEL classifications: G10, G13, G17.

1. Introduction

In financial markets, time variation in discount rates has proven to be the driving force behind price fluctuations. Cochrane, in his presidential address (Cochrane 2011), observes that the pervasive feature of asset pricing is that empirically, prices forecast future discount rates rather than future cash flows. This pattern holds even for durable goods markets such as housing. An open question is whether this same logic extends to commodities markets.

In this article, we show that the commodities market is in fact quite different. Commodity prices can still be viewed as reflecting the discounted value of sequences of future cash flows; however, unlike other markets, prices strongly predict cash flows. This makes commodity prices much closer to the classical textbook view of price changes reflecting cash flow news.

From the point of view of a consumer of a commodity, at first glance it may seem peculiar to talk about the role of discount rates versus cash flows. If you operate an airline, a barrel of oil is purely a one-time cash outlay, and the only question is whether to set the price now on the futures market or later on the spot market. From the point of view of an *investor*, though, the commodity plays a different role. An investor can continually trade in the spot and futures markets to make a profit on the difference between spot and futures prices. Spot and futures prices do not move together perfectly, and the difference can be a reflection of news about the profitability of investing in this market, or how heavily

Received: June 7, 2024. **Accepted:** November 5, 2024
Editor: Ian Dew-Becker

investors discount. Experience in other markets suggests that the discount rate should dominate. It does not.

To get at this question requires several steps. As a first step, a single unit of a commodity can be turned into an infinite series of cash flows by a simple trading strategy. Each period, the investor sells the unit and enters into a forward contract to buy the unit next period. The investor then buys a bond today to cover the expense of the forward later. The difference between the sale price of the commodity and the cost of the bond is a cash flow reflecting the *net convenience yield* of the commodity, the convenience yield net of storage costs.[1] The investor can repeat this strategy next period and indefinitely thereafter, generating an infinite sequence of net convenience yields (this observation goes back to Pindyck 1993). By the logic of net present value, the original commodity price must equal the discounted sum of these net convenience yields. Moreover, changes in prices should reflect either changes in expected yields, discount rates, or both.

The next step is to linearize the present-value relationship between prices and cash flows. Traditionally, the Campbell and Shiller (1988) decomposition is used for this purpose. For stocks, this model implies that the dividend–price ratio can *only* vary if it forecasts expected returns, dividend growth, or a potential bubble component. The accumulated evidence suggests that most of the predictability comes from expected return and not the other two components (e.g., Cochrane 2008, 2011). Such predictability also prevails in various other asset markets, that is, corporate bond yields forecast bond returns, and rent-house price ratios forecast housing returns (e.g., Cochrane 2011).[2]

Extending the linearized present-value model to commodity prices is not completely straightforward. Since the net convenience yield represents the cash flow (payoff) of commodities, the analog of the dividend–price ratio is the percentage net convenience yield, which is the net convenience yield divided by the commodity price. One significant difference from the stock case is that the net convenience yield can be negative. If the forward price is significantly higher than the spot price, then the implied cash flow is negative. While this does not cause any trouble for the present-value interpretation, it does make it difficult to apply the Campbell–Shiller decomposition, which requires the dividend–price ratio to be positive to apply log-linearization. Pindyck (1993) offers a partial solution, namely by linearizing the return definition directly, without taking logs. Pindyck's approach further assumes that net convenience yields are discounted with a fixed risk premium, which forecloses the possibility of testing whether prices predict changes in that risk premium. The approximation also has large errors, particularly for negative net convenience yields, which results in errors in assessing the predictability of future discount rates and future cash flows.

To solve this problem, we pursue an alternative approach. We adopt the *neglog* transformation (Whittaker, Whitehead, and Somers 2005) to generalize the Campbell–Shiller linearized return identity for negative yields. The neglog transform is particularly useful for financial variables and allows us to map the effects of compounding on a linear scale when yields are negative, while preserving the *economic* interpretation of the variables of interest. That is, irrespective of whether yields are positive or negative, a higher price relates to a lower (neglog) percentage yield (percentage net convenience yield), a higher net convenience yield leads to a higher (neglog) percentage yield, and a higher future net convenience yield relates to larger (neglog) yield growth (net convenience yield growth). Moreover, the (neglog) yield growth rates remain easily compoundable through straightforward summation.

With the generalized linear return identity at hand, we can extend the logic of the Campbell–Shiller decomposition to commodities. We relate the volatility of the percentage yields to the predictability of commodity returns, yield growth, and future percentage

[1] To be sure, this is not simply the "non-pecuniary benefit" of holding a commodity as the net convenience yields is usually described; this strategy generates a monetary cash flow for the investor.

[2] In this literature, predictability of expected returns is interpreted as a reflection of discount rate variation, so we use expected returns and discount rates interchangeably.

yields. The percentage yield can move at all only if it forecasts returns, yield growth, or percentage yields, or some combination of the three. We assess these relations for several investment horizons and obtain the implied long-run relationships from a vector autoregressive (VAR) analysis. As in Cochrane's (2008) study, we also conduct a Monte Carlo simulation to *jointly* test and explain the predictability of commodity returns and yield growth based on the decomposition.

Using a commodity index based on a bimonthly sample of twenty-three commodities for the period 1959–2024, we observe coherent predictability patterns in both commodity returns and yield growth. Higher commodity prices are associated with higher expected net convenience yields, as well as lower expected returns. Notably, the predictability of yield growth is stronger than that of commodity returns, which we link to the strong correlation between yield growth shocks and "bubble" shocks. In the long run, variation in commodity prices appears predominantly driven by time-varying expected yield growth. These findings are insensitive to the method of commodity index construction and seasonality characteristics of commodities. Moreover, 4-year horizon regressions using only positive yields and the standard Campbell–Shiller decomposition indicate that the results are not driven by the neglog transformation.

Our article contributes to the literature on the determinants of commodity prices. Two popular views of commodity prices are elaborated in seminal work by Fama and French (1987): (1) the storage theory (e.g., Casassus and Collin-Dufresne 2005; Ye, Zyren, and Shore 2005; Brooks, Prokopczuk, and Wu 2013; Fernandez 2020; Jacks and Stuermer 2020), linking futures prices to interest rates, storage costs, and convenience yields and (2) the risk premium theory (e.g., Fama and French 1987; Brooks, Prokopczuk, and Wu 2013; Narayan, Narayan, and Sharma 2013), decomposing commodity futures prices into expected risk premiums and future spot price forecasts. These theories suggest a close relationship between spot and futures prices, with various factors such as basis, convenience yield, and demand shocks influencing commodity prices. The risk-premium approach is most closely related to ours in the sense that it also analyzes a particular decomposition for commodity prices.

In addition to commodity fundamentals, macroeconomic and financial variables such as inflation, interest rates, and equity premiums also impact commodity returns (e.g., Chen, Rogoff, and Rossi 2010; Gargano and Timmermann 2014; Lutzenberger 2014; Kagraoka 2016; Watugala 2019). Another strand of literature delves into the characteristics of convenience yields, indicating its relationship with inventory levels and production costs (e.g., Milonas and Henker 2001; Kocagil 2004; Mirantes, Población, and Serna 2013).

The literature also points out that commodities have become an increasingly important asset class as an additional source of diversification (e.g., Belousova and Dorfleitner 2012; van Huellen 2019). This importance is reflected by an unprecedented capital inflow from investors into commodity markets, a phenomenon known as "the financialization of commodities" (e.g., Tang and Xiong 2012; Basak and Pavlova 2016; van Huellen 2019). Consequently, commodities exhibit characteristics akin to conventional assets (e.g., Domanski and Heath 2007).

In response, commodities have gained notable attention in the recent finance literature (e.g., Bakas and Triantafyllou 2018; Prokopczuk, Stancu, and Symeonidis 2019; Kang, Rouwenhorst, and Tang 2020; Christoffersen, Jacobs, and Pan 2022; Gao et al. 2022; Goldstein and Yang 2022; Patton and Weller 2022; Ready and Ready 2022; Elkamhi and Jo 2023; Han 2023; Hazelkorn, Moskowitz, and Vasudevan 2023). However, the discount rate connection is still unexplored, namely whether commodity prices forecast discount rates, expected cash flows, or both. To our knowledge, we are the first to explicitly address this question, and our findings suggest it is expected cash flows.

The rest of this article is organized as follows. In Section 2, we present the method to calculate the net convenience yield, the generalized Campbell–Shiller return decomposition, VAR regressions, and Monte Carlo simulations. In Section 3, we describe the data. We report the results in Section 4 and conduct robustness checks in Section 5. Section 6 concludes.

2. Methodology

2.1 The net convenience yield and commodity returns

We begin by decomposing the spot price as a discounted sum of net convenience yields. Consider an investor who owns a commodity and engages in the following strategy. At time t, the investor sells the commodity at the spot price S_t and invests this amount in a risk-free asset, at the one-period risk-free interest rate, $rf_{t \to t+1}$ (rf_t). At the same time, the investor engages in a futures contract, promising to buy back the commodity at the futures price $F_{t,t+1}(F_t)$ at $t+1$. As is a common practice, the futures price is determined in the market in such a way that neither buyer nor seller of the contract needs to be compensated. The resulting payoff of this strategy is not necessarily zero, since during this period the counterparty gains a latent payoff (the convenience yield) by holding the commodity but also faces storage costs, and the difference between the futures and spot prices reflects this. The payoff of the strategy is thus equal to the net convenience yield, that is, the convenience yield net of storage cost.[3] It *implicitly* measures the net economic benefit of physically holding a commodity (i.e., the ability to avoid stockouts, tackle an unexpected production arrangement, etc.) compared to holding a claim to that same commodity at a future date.

Let $D^t_{t \to t+1}$ be the net convenience yield from time t to time $t+1$,

$$D^t_{t \to t+1} = S_t - \frac{F_t}{1 + rf_t}. \tag{1}$$

Note that the net convenience yield is known at time t and can also be inferred exclusively from futures prices.[4] At $t+1$, the investor can repeat the strategy, and as such, an owner of a commodity at time zero can turn this into an infinite stream of one-period cash flows, $D^t_{t \to t+1}, \forall t$. The price of the commodity should therefore be equal to the present value of the associated expected cash flows, $D^t_{t \to t+1}$ (Pindyck 1993). This analogy between commodity prices and net convenience yields, on the one hand, and stock prices and dividends, on the other hand, enables us to extend the Campbell–Shiller analysis to this class of investments. However, this approach adopts a log-linearized *return* equation to analyze the price variation implications of the net present-value equation.

Because the investor is always *implicitly* compensated by the net convenience yield, the appropriate definition of a commodity return obtained at time $t+1$ should include the net convenience yield, instead of only looking at capital gains (price changes) (e.g., Narayan, Narayan, and Sharma 2013; Wang, Liu, and Wu 2020). So if the net convenience yield from time t to time $t+1$ is obtained at time t, the realized return by holding a commodity in this period is defined as:

$$R_{t+1} \equiv \frac{S_{t+1}}{S_t - D_t} \equiv \frac{F_{t+1}/(1 + rf_{t+1}) + D_{t+1}}{F_t/(1 + rf_t)} \equiv \frac{\tilde{F}_{t+1} + D_{t+1}}{\tilde{F}_t}, \tag{2}$$

where $\tilde{F}_t = \frac{F_t}{1+rf_t}$ and $D_t = S_t - \frac{F_t}{1+rf_t} = S_t - \tilde{F}_t$. S_t is the spot price at time t, F_t is the futures price at time t, at which the commodity will be delivered at the end of the period, at time

[3] The net convenience yield $D^t_{t \to t+1}$ is different from the "basis." Basis is defined as the difference of contemporaneous futures and spot prices, that is, $F_t - S_t$. So a high net convenience yield implies a low basis, as predicted by the theory of storage.

[4] We highlight how $D^t_{t \to t+1}$ can also be inferred exclusively from futures prices in the Supplementary material.

$t+1$. rf_t is the one-period risk-free rate from time t to $t+1$.[5] Note that unlike dividends, $D^t_{t\to t+1}$ can be both positive and negative.[6] The possibility of negative yields is not problematic for the definition of returns, but it complicates the application of the Campbell–Shiller decomposition, an issue we address in the next section.

2.2 Generalized Campbell–Shiller return decomposition for negative yields

In this section, we generalize the Campbell–Shiller decomposition so that it can also be applied with negative net convenience yields. We would like to obtain an identity among returns, net convenience yield growth, and the net percentage yield; however, we cannot apply the standard Campbell and Shiller (2015) log-linearized return decomposition, since D_t can be negative. To resolve this issue, we adopt the *neglog* transformation, which allows us to deal with negative net convenience yields while preserving the economic interpretation of the variables of interest.

2.2.1 The neglog transformation

The neglog transformation of variable x is defined as:

$$\mathrm{nl}_\alpha(x) := \alpha \cdot \mathrm{sgn}(x) \cdot \ln\left(\frac{|x|}{\alpha} + 1\right), \tag{3}$$

with the constant $\alpha > 0$ a normalization parameter and for the standard neglog $\alpha = 1$ (Whittaker, Whitehead, and Somers 2005). The neglog function passes through the origin, is monotonically increasing and continuous, and its first derivative is also continuous at the origin.[7] It has similar advantages to the natural logarithm and extends monotonically for negative values. Moreover, the effects of compounding are transformed to a linear scale, while relative changes on this scale facilitate economic interpretation, that is, the signs of the relative changes correspond to gains and losses (Whittaker, Whitehead, and Somers 2005).[8]

The neglog transformation is particularly useful for financial variables, for example, interest rates and percentage yields, which are usually relatively "small" numbers so that they are near the origin of the neglog transformation. However, the level of the net convenience yield is not necessarily "close to the origin" and its (absolute) value tends to increase over time. This can disrupt some of the attractive properties of the neglog transformation, but we can circumvent this issue by working with detrended variables.

2.2.2 Defining the variables of interest using the neglog transformation

Without loss of generality, we define detrended variables $\hat{D}_t = D_t/(D_0 e^{\mu t})$ and $\hat{F}_t = \tilde{F}_t/(F_0 e^{\mu t})$, where the constants $D_0, F_0 > 0$ are initial values, and we pick the constant μ equal to the average growth rate of $\tilde{F}_t$ for $t = 0, .., T$ (i.e., $\mu = \ln(\tilde{F}_T/\tilde{F}_0)/T$). Our

5 Equation (2) is equivalent to the following present-value identity: $\tilde{F}_t = R^{-1}_{t+1}(D_{t+1} + \tilde{F}_{t+1}) = R^{-1}_{t+1}D_{t+1} + R^{-1}_{t+1}R^{-1}_{t+2}D_{t+2} + \cdots = \sum_{j=1}^{\infty}(\prod_{k=1}^{j} R^{-1}_{t+k})D_{t+k} = R^{-1}_{t+1}(D_{t+1} + \sum_{j=2}^{\infty}(\prod_{k=2}^{j} R^{-1}_{t+k})D_{t+k}) = R^{-1}_{t+1}S_{t+1}$. Recall that the net convenience yield is the latent payoff of holding a commodity and can be collected by holding a futures contract. Actually, the variations in futures prices $\tilde{F}_t$ and spot prices S_t are two sides of the same coin.

6 The reason for this is that for commodities there can be both substantial storage costs and convenience yields. When $D^t_{t\to t+1}$ is negative, it implies that the investor pays less for the commodity relative to engaging in a futures contract of the commodity. If investors own the commodity, they have to store it, but with a futures contract they do not. The monetary amount $D^t_{t\to t+1}$ then reflects the compensation for the storage cost, that is, it represents the amount the commodity investor requires from the owner of a futures contract to compensate for the storage cost. If $D^t_{t\to t+1}$ is positive, this is driven by the relatively large convenience yield. For example, if the investors suddenly physically need the commodity between time t and $t+1$—such as oil during an energy crisis—it is advantageous for them to currently own the physical commodity. So the value reflects the monetary amount that they are willing to pay to the owner of a futures contracts to enjoy the convenience yield net of the storage cost.

7 The neglog takes inspiration from the broader class of power transformations introduced by Yeo and Johnson (2000); for a visualization, we refer to their original article.

8 For example, a change in net convenience yields from $D_t = -5$ to $D_{t+1} = 10$ is clearly an economic gain, yet a standard gross growth rate calculation gives $D_{t+1}/D_t = 10/(-5) = -200\%$, which reads as a loss.

variables of interest (not yet in logs), that is, gross return (R_t), gross yield growth (ΔD_t), and percentage yield (Y_t), can be expressed in terms of these detrended variables:

$$R_{t+1} \equiv \frac{\tilde{F}_{t+1} + D_{t+1}}{\tilde{F}_t} = \frac{\hat{F}_{t+1} + e^{y_0}\hat{D}_{t+1}}{\hat{F}_t} \cdot e^{\mu}, \tag{4}$$

$$\Delta D_{t+1} \equiv D_{t+1}/D_t = \hat{D}_{t+1}/\hat{D}_t \cdot e^{\mu}, \tag{5}$$

$$Y_t \equiv D_t/\tilde{F}_t = \hat{D}_t/\hat{F}_t \cdot e^{y_0}, \tag{6}$$

where $y_0 \equiv \ln(D_0/F_0)$.

The standard Campbell–Shiller return identity derivation proceeds with taking logs of these variables, which in terms of our detrended variables would give: $\ln(R_{t+1}) = \ln(\hat{F}_{t+1} + e^{y_0}\hat{D}_{t+1}) - \ln(\hat{F}_t) + \mu$, $\ln(\Delta D_{t+1}) = \ln(\hat{D}_{t+1}) - \ln(\hat{D}_t) + \mu$, and $\ln(Y_t) = \ln(\hat{D}_t) - \ln(\hat{F}_t) + y_0$. However, the net convenience yield, D_t, can be negative, so we cannot take logs for the variables of interest. Note that taking logs is not problematic for the returns (since $R_{t+1} > 0, \forall t$), but it is problematic for the convenience yield growth and percentage yield. Instead, we adopt analogue formulas for our variable definitions, replacing the log with the neglog for $\hat{D}_t$:

$$r_{t+1} := \ln(R_{t+1}), \tag{7}$$

$$\Delta d_{t+1} = \mathrm{nl}_\alpha(\hat{D}_{t+1}) - \mathrm{nl}_\alpha(\hat{D}_t) + \mu, \tag{8}$$

$$y_t := \mathrm{nl}_\alpha(\hat{D}_t) - \ln(\hat{F}_t) + y_0. \tag{9}$$

We interpret the variables r_{t+1} as the log return, Δd_{t+1} as the neglog net convenience yield growth, and y_t as the neglog net percentage convenience yield. We often refer in short to the three variables as commodity return, yield growth and percentage yield. Here, the neglog net percentage convenience yield y_t preserves the properties that a lower percentage yield relates to a lower net convenience yield or a higher price ($\frac{\partial y_t}{\partial D_t} > 0, \forall D_t, \tilde{F}_t$ and $\frac{\partial y_t}{\partial \tilde{F}_t} < 0, \forall D_t, \tilde{F}_t$). Furthermore, a higher yield growth relates to a higher future net convenience yield or a lower current net convenience yield ($\frac{\partial \Delta d_{t+1}}{\partial D_{t+1}} > 0$ and $\frac{\partial \Delta d_{t+1}}{\partial D_{t+1}} > 0$), and compounding growth is achieved by simple summation,[9] that is, $\Delta d_{t \to t+k} := \mathrm{nl}_\alpha(\hat{D}_{t+k}) - \mathrm{nl}_\alpha(\hat{D}_t) + k\mu = \sum_{i=1}^{k} \Delta d_{t+i}$. Moreover, higher prices mean lower net percentage yields and an increase in net convenience yields implies larger net yield growth, even if yields are negative.

2.2.3 The generalized Campbell–Shiller return identity

Next, we rewrite equation (2) into an approximate linear identity in terms of the variables r_{t+1}, Δd_{t+1}, and y_t, by first taking the natural log of the return definition in terms of the detrended variables, and then performing a Taylor series expansion of the variables $\ln(\hat{F}_{t+1})$ and $\mathrm{nl}_\alpha(\hat{F}_{t+1})$ around the points $\hat{F}_{t+1} = 1$ and $\hat{D}_{t+1} = 1$:[10]

[9] The formula for the growth rates basically uses the neglog to assess in percentages how much D_{t+1} and D_t deviate from their trend, takes the difference between these numbers, and adds back the average growth rate. Suppose $T = 3$ and $D_0 = 2, D_1 = -2, D_2 = -1, D_3 = 4$. We have for the average growth rate $\mu = \ln(D_3/D_0)/3 = 23\%$. The neglog growth rates with $\alpha = 1$ are $\Delta d_1 = -105\%$, $\Delta d_2 = 54\%$, and $\Delta d_3 = 120\%$, so positive rates are associated with increases in D_t and negative rates with decreases in D_t. Moreover, $d_1 + d_2 + d_3 = 69\% = 3\mu$. The calculation of the net percentage yield is in the same spirit.

[10] We justify linearization around these points by implicitly assuming that D_t and $\tilde{F}_t$ are trend stationary so that $\hat{F}_t$ and $\hat{D}_t$ remain close to 1. The diagnostics to empirically assess whether this is the case are described in Section 3.2.

$$\begin{aligned}
\ln(R_{t+1}) &\equiv \ln(\hat{F}_{t+1} + e^{y_0}\hat{D}_{t+1}) - \ln(\hat{F}_t) + \mu \\
&= \ln\left(\hat{F}_{t+1} + \alpha e^{y_0}\operatorname{sgn}(\hat{D}_{t+1})\left(\frac{|\hat{D}_{t+1}|}{\alpha} + 1 - 1\right)\right) - \ln(\hat{F}_t) + \mu \\
&= \ln\left(e^{\ln(\hat{F}_{t+1})} + \alpha e^{y_0}\operatorname{sgn}(\hat{D}_{t+1})\left(e^{\mathrm{nl}_\alpha(\hat{D}_{t+1})/(\alpha\operatorname{sgn}(\hat{D}_{t+1}))} - 1\right)\right) - \ln(\hat{F}_t) + \mu \qquad (10) \\
&\approx \ln(1 + e^{y_0}) + \frac{1}{1+e^{y_0}}\left(\ln(\hat{F}_{t+1})\right) + \frac{e^{y_0}}{1+e^{y_0}}\left(\mathrm{nl}_\alpha(\hat{D}_{t+1}) - \mathrm{nl}_\alpha(1)\right) \\
&\quad - \ln(\hat{F}_t) + \mu.
\end{aligned}$$

Defining the constant of linearization $\rho = \frac{1}{1+e^{y_0}}$ and adding and subtracting the term $\mathrm{nl}_\alpha(\hat{D}_t) + (1-\rho)y_0$, we can rewrite equation (10) as:

$$\begin{aligned}
\ln(R_{t+1}) &= \kappa - \rho[\mathrm{nl}_\alpha(\hat{D}_{t+1}) - \ln(\hat{F}_{t+1}) + y_0] + [\mathrm{nl}_\alpha(\hat{D}_{t+1}) - \mathrm{nl}_\alpha(\hat{D}_t) + \mu] \\
&\quad + [\mathrm{nl}_\alpha(\hat{D}_t) - \ln(\hat{F}_t) + y_0],
\end{aligned} \qquad (11)$$

where the constant $\kappa = \ln(1 + e^{y_0}) - (1-\rho)(\mathrm{nl}_\alpha(1) + y_0)$. This equation can be rewritten in terms of our variables of interest as:

$$r_{t+1} = \kappa - \rho y_{t+1} + \Delta d_{t+1} + y_t. \qquad (12)$$

We thus preserve the economic interpretation of the standard Campbell and Shiller (2015) return decomposition related to equation (12): Returns are high when prices increase ($y_{t+1}\downarrow$), or if the net convenience yields increase ($\Delta d_{t+1}\uparrow$) with no change in net percentage convenience yields ($y_{t+1} = y_t$). In fact, if $D_t > 0, \forall t$, and we set $\alpha = 1$ and $\mu = 0$, then as $\lim D_0 \to 0$ we observe that in the limit the standard Campbell and Shiller (2015) decomposition is a special case of equation (12), with the standard definitions of log yields and log growth.[11]

2.3 VAR regressions

We proceed as in Cochrane (2008) for the standard Campbell–Shiller case and rearrange equation (12) as:

$$y_t = r_{t+1} - \Delta d_{t+1} + \rho y_{t+1} - \kappa, \qquad (13)$$

which holds both *ex post* but also *ex ante*, so we can take the expectation of both sides of the equation at time t. This implies that the volatility of the current net percentage convenience yield, y_t, is related to the variation in the expectation of the commodity return, the expectation of yield growth, and/or the expectation of the next-period net percentage convenience yield.

[11] In this case, $\Delta d_{t+1} = \mathrm{nl}_\alpha(\hat{D}_{t+1}) - \mathrm{nl}_\alpha(\hat{D}_t) = \operatorname{sgn}(D_{t+1})\ln\left(\frac{|D_{t+1}|}{D_0} + 1\right) - \operatorname{sgn}(D_t)\ln\left(\frac{|D_t|}{D_0} + 1\right) = (+)\ln(D_{t+1} + D_0) - (+)\ln(D_t + D_0)$, which implies $\Delta d_{t+1} \to \Delta\ln(D_{t+1})$ as $D_0 \to 0$. Furthermore, $y_t = \mathrm{nl}_\alpha(\hat{D}_t) - \ln(\hat{F}_t) + y_0 = (+)\ln(D_t \;\; D_0 + 1) - \ln(\tilde{F}_t) - \ln(F_0) + \ln(D_0) + \ln(F_0) = \ln(D_t + D_0) - \;\; \ln(\tilde{F}_t)$, so $y_t \to \ln\left(\frac{D_t}{\tilde{F}_t}\right)$ as $D_0 \to 0$. Hence, for strictly positive variables, the limiting case gives the standard definitions for log growth and log yields.

We can run the one-period VAR regressions to determine which component dominates:

$$\begin{aligned} r_{t+1} &= a_r + \beta_r y_t + \varepsilon^r_{t+1}, \\ \Delta d_{t+1} &= a_d + \beta_d y_t + \varepsilon^d_{t+1}, \\ y_{t+1} &= a_y + \beta_y y_t + \varepsilon^y_{t+1}. \end{aligned} \tag{14}$$

With the restriction implied by equation (13), an approximate identity linking the regression coefficients is[12]:

$$\beta_r - \beta_d + \rho\beta_y = 1. \tag{15}$$

So identity (15) implies that[13]

$$var(y_t) = cov(r_{t+1}, y_t) - cov(\Delta d_{t+1}, y_t) + \rho cov(y_{t+1}, y_t). \tag{16}$$

Therefore, the percentage yield, y_t, can *only* vary if it forecasts commodity returns, future yield growth, or future percentage yields, or some combination of the three. The coefficients β_r, $-\beta_d$, and $\rho\beta_y$ represent the contribution of the three components in driving the movement of y_t for a one-period horizon.

The coefficient β_y cannot be too large, and $\rho\beta_y$ should be smaller than one. Otherwise, the percentage yield is explosive, which is obviously not economically plausible, as argued by Cochrane (2008) for the case of stock dividend yields. Identity (15) then implies that $\beta_r = 0$ and $\beta_d = 0$ cannot simultaneously be true if there is variation in net percentage yields. Therefore, the percentage yield *must* forecast either future commodity return or future yield growth or both. We iterate equation (13) and obtain a present-value identity:

$$y_t = \sum_{j=1}^{h} r_{t+j}\rho^{j-1} - \sum_{j=1}^{h} \Delta d_{t+j}\rho^{j-1} + \rho^h y_{t+h} - k\frac{1-\rho^h}{1-\rho}, \tag{17}$$

where h refers to the length of horizon, $\sum_{j=1}^{h} r_{t+j}\rho^{j-1}$ is the weighted h-period commodity return, and $\sum_{j=1}^{h} \Delta d_{t+j}\rho^{j-1}$ is the weighted h-period yield growth. Again, yields should predict long-run returns, long-run yield growth, or a "bubble". Long-horizon coefficients can be estimated directly from:

$$\begin{aligned} \sum_{j=1}^{h} r_{t+j}\rho^{j-1} &= a_r^{(h)} + \beta_r^{(h)} y_t + \varepsilon^r_{t+h}, \\ \sum_{j=1}^{h} \Delta d_{t+j}\rho^{j-1} &= a_d^{(h)} + \beta_d^{(h)} y_t + \varepsilon^d_{t+h}, \\ y_{t+h} &= a_y^{(h)} + \beta_y^{(h)} y_t + \varepsilon^y_{t+h}. \end{aligned} \tag{18}$$

The long-horizon coefficients should satisfy the restriction $\beta_r^{(h)} - \beta_d^{(h)} + \rho^h\beta_y^{(h)} = 1$. The long-horizon coefficients can also be inferred from the one-period horizon coefficients, by $\beta_r^{(h)} = \frac{\beta_r(1-(\rho\beta_y)^h)}{1-\rho\beta_y}$, $\beta_d^{(h)} = \frac{\beta_d(1-(\rho\beta_y)^h)}{1-\rho\beta_y}$, and $\beta_y^{(h)} = (\beta_y)^h$.

Explosive yields are ruled out, so $\rho\beta_y < 1$. In the long run, when h approaches infinity, the contribution of percentage yields vanishes. Therefore, in the long run, all volatility in

12 The error terms should also meet the restriction that $\varepsilon^r_{t+1} - \varepsilon^d_{t+1} + \rho\varepsilon^y_{t+1} = 0$.

13 Because regression coefficients are defined as: $\beta_r = \frac{cov(r_{t+1}, y_t)}{var(y_t)}$, $\beta_d = \frac{cov(\Delta d_{t+1}, y_t)}{var(y_t)}$, $\beta_y = \frac{cov(y_{t+1}, y_t)}{var(y_t)}$.

percentage yields comes from the predictability of commodity returns and/or yield growth. The long-run coefficient for returns $\beta_r^{(lr)}$ in terms of the one-period horizon coefficient when h goes to infinity is then simplified, namely $\beta_r^{lr} = \frac{\beta_r}{1-\rho\beta_y}$. Similarly, the long-run coefficient for yield growth is then equal to $\beta_d^{lr} = \frac{\beta_d}{1-\rho\beta_y}$. The coefficients β_r^{lr} and $-\beta_d^{lr}$ represent their contributions in driving the variation of percentage yields in the long run and hence provide an answer to which of the two drives variation in scaled commodity prices.

2.4 Monte Carlo simulations for joint testing

As implied by the coefficient identity (15), if returns are not predictable, yield growth must be predictable, and vice versa. Therefore, it is better to jointly study the predictability of commodity returns and yield growth, as argued by Cochrane (2008). In this section, we introduce Monte Carlo simulations based on the one-period VAR estimates to jointly test the predictability.

We start with the estimated covariance matrix of error terms in equation (14). We generate normally distributed error terms ε^d and ε^y and obtain the implied ε^r from the implied identity of for the errors. We generate the starting values of percentage yields from the unconditional density $y_0 \sim N[0, \sigma^2(\varepsilon^y)/(1-\beta_y^2)]$.[14] Following Cochrane (2008), we first test the null hypothesis that returns are not predictable and yield growth is predictable. $H_0: \beta_r = 0, \beta_d = \rho\beta_y - 1$ or $H_0: \beta_r^{lr} = 0, -\beta_d^{lr} = 1$. The coefficients are restricted with identity (15). This null hypothesis implies that

$$\begin{bmatrix} r_{t+1} \\ \Delta d_{t+1} \\ y_{t+1} \end{bmatrix} = \begin{bmatrix} 0 \\ \rho\beta_y - 1 \\ \beta_y \end{bmatrix} y_t + \begin{bmatrix} \varepsilon_{t+1}^d - \rho\varepsilon_{t+1}^y \\ \varepsilon_{t+1}^d \\ \varepsilon_{t+1}^y \end{bmatrix}. \tag{19}$$

We also test an alternative hypothesis that the returns are predictable and the yield growth is not predictable. $H_0: \beta_r = 1 - \rho\beta_y, \beta_d = 0$ or $H_0: \beta_r^{lr} = 1, \beta_d^{lr} = 0$. This null hypothesis implies that

$$\begin{bmatrix} r_{t+1} \\ \Delta d_{t+1} \\ y_{t+1} \end{bmatrix} = \begin{bmatrix} 1 - \rho\beta_y \\ 0 \\ \beta_y \end{bmatrix} y_t + \begin{bmatrix} \varepsilon_{t+1}^d - \rho\varepsilon_{t+1}^y \\ \varepsilon_{t+1}^d \\ \varepsilon_{t+1}^y \end{bmatrix}. \tag{20}$$

We generate the simulated return, yield growth, and percentage yield according to equation (19) under the first null hypothesis and according to equation (20) under the second null hypothesis. After we have all the simulated data (note that all the simulated data are demeaned), we run the VAR regression (18) at different horizons and collect the simulated coefficients. This process is repeated 50,000 times matching the length of the sample period, following Golez and Koudijs (2018). The distribution of the simulated coefficients helps to understand the probability that we by chance obtain the return or yield growth coefficient estimated from real commodity data if the return or yield growth is not predictable.

3. Data

3.1 Data sources

We collect end-of-month close prices from the Commodity Research Bureau (CRB) to calculate month-end to month-end returns and yield growth in line with Szymanowska et al. (2014) and Boons and Prado (2019). As is common practice in the commodity literature,

[14] Since $y_{t+1} = \beta_y y_t + \varepsilon_{t+1}^y$, so $\sigma^2(y) = \beta_y^2\sigma^2(y) + \sigma^2(\varepsilon^y)$ and $\sigma^2(y) = \sigma^2(\varepsilon^y)/(1-\beta_y^2)$.

the nearest-to-maturity (nearby) futures price is used to proxy for the spot price because of illiquid markets (e.g., Kocagil 2004; Symeonidis et al. 2012; Brooks, Prokopczuk, and Wu 2013; Yang 2013). A general way to generate the nearest-to-maturity futures prices is to roll over to the next nearest-to-maturity futures contract at the end of the month just before the delivery month of the nearest-to-maturity futures contract. However, erratic price and volume behavior can occur for futures contracts during both the delivery month and the month prior because investors start to roll over to the next futures contract from 4 to 6 weeks before the nearest-to-maturity futures contract expires (e.g., Brunetti and Reiffen 2014). To prevent this from affecting the results, we roll over to the next futures contract two months rather than one month before the delivery months to proxy for spot prices, following He, Jiang, and Molyboga (2019) and Szymanowska et al. (2014). The futures prices of contracts maturing in 2 and 4 months are proxies for spot prices and futures prices maturing 2 months later. For instance, in January, we use the price of futures contract expiring in March to proxy for the spot price, and use the price of a futures contract expiring in May to proxy for the futures price of a contract maturing in March.

Additionally, not all commodities have futures contracts that mature in every month. For example, the futures contracts written on soybeans only have delivery dates in January, March, May, July, August, September, and November. In order to minimize the problem raised by the irregular delivery dates, we select the twenty-three commodities from July 1959 to January 2024 whose futures contracts have delivery dates that are spread out as evenly as possible for every 2 months and construct spot and futures prices at the *bi-monthly* frequency, in line with, for example, Han (2023) and Szymanowska et al. (2014).[15] We use the set of twenty-one commodities from Szymanowska et al. (2014), to which we add gasoil and natural gas, as in Boons and Prado (2019). Using futures prices to proxy spot prices allows us to use high-quality data and only has a small effect on the results, even from a *theoretical* point of view, which is shown and argued in detail in the Supplementary material. More details about the twenty-three commodities are shown in Appendix Table B1.

3.2 Commodity indices, variable construction, and validation

We create two commodity indices, a scaled price-weighted index and an equally weighted index. Value-weighted indices are most common for stocks, but for commodities it is not so clear which amounts should be used as quantities to multiply with the price to determine weights. To stay as close as possible to the intuition of value weighting, we instead construct a price-weighted commodity index. In order to avoid the concern that the index is driven by few commodities with high prices, we use *scaled* commodity prices by rescaling them with their time-series average value to determine their corresponding weights. With the scaled price-weight, we define the commodity spot index as $S_{t+1} = \Delta S_{t+1} S_t$, where $\Delta S_{t+1} = \sum_{i=1}^{N} w_{i,t} \Delta S_{i,t+1}$, $\Delta S_{i,t+1} = S_{i,t+1}/S_{i,t}$, $w_{i,t} = S^s_{i,t} / \sum_{i=1}^{N} S^s_{i,t}$, and $S^s_{i,t}$ is the scaled price of commodity i, and N is the number of commodities in the portfolio. We refer to it below as the price-weighted index. Accordingly, we define the futures index as $F_t = \sum_{i=1}^{N} n_{i,t} F_{i,t}$, where $n_{i,t} = w_{i,t} V_t / S_{i,t}$, and $V_t = S_t$ in order to make the gross net convenience yield of the index consistent. In addition, we create an equally weighted index, which is more common

[15] Some commodities have futures contracts that expire exactly every 2 months (January, March, May, July, September, and November), for example, soybean, while some commodities do not. For these commodities, we use the price of the futures contract with the nearest maturity after 2 and 4 months to proxy spot prices and futures prices of contracts maturing 2 months later. For example, cocoa has futures contracts that expire in March, May, July, September, and December. The spot prices of cocoa for September and November are represented with the futures prices of contracts maturing in December and March instead of November and January. Some commodities, for example, gold and live cattle, have futures contracts expiring in February, April, June, August, October, and December. For these commodities, we use the price of futures contracts maturing in 3 months to proxy for spot prices. For instance, we use the price of futures contracts maturing in April to proxy spot prices in January.

in the commodity literature. The most important difference between the two indices is that the price-weighted index represents a buy-and-hold strategy, whereas the equally weighted index requires rebalancing. More details about the index construction are explained in Appendix A.1.[16]

A risk-free interest rate is also required to calculate the net convenience yield. We use the 1-month Treasury bill rates from Kenneth French's website. We then compound the monthly interest rate to a 2-month interest rate at bimonthly frequency, in line with the definition of rf_t in Section 2.1. Our full sample ranges from July 1959 to January 2024. We use annual return, yield growth, and percentage yield at *bimonthly* frequency in the analysis.[17]

To perform the neglog-linearization, we must pick values for the normalization parameters α, μ, F_0, and D_0. As this choice is somewhat arbitrary[18]—note that the linearization works for any positive value of these parameters—we explain our choice in Appendix A.2. As a validation check, we compare with the results when applying the same transformation to stock market data, as we discuss below.

Ultimately, the most important consideration in the choices for α, μ, F_0, and D_0 is that $\hat{F}_t$ and $\hat{D}_t$ remain "close" to 1 so that the linearization around 1 indeed yields an approximate identity. In light of this, we also empirically check the accuracy of the linearization in our subsequent analysis by assessing whether the regression coefficients indeed meet the implied restriction in equation (13). Moreover, when determining the value for the parameter D_0 using GMM, we focus on matching the variances of the neglog percentage yield and neglog yield growth with the associated variances of the standard log percentage yield and log yield growth (based on the subsample for which the convenience yield is positive), since the goal of this article is to assess how much of the variance in the percentage yield can be explained by the variance of the other variables of interest.[19] This way, we are confident that we are comparing variables on similar scales.

In order to validate the use of the neglog-linearization, we apply the same transformation to standard US stock market data. The summary statistics of the neglog dividend growth, neglog dividend yield, and their corresponding standard log variables for US stocks are shown in Appendix Table B2. The variances of the neglog dividend growth and neglog dividend yield are similar to those of their corresponding log variables. Scatter plots in Appendix figure C1 show a positive and monotonic relationship between the neglog dividend growth (neglog dividend yield) and the standard log dividend growth (log dividend yield). As such, this suggests that neglog-linearization is a suitable substitute for log-linearization when cash flows can be negative.

3.3 Data description

Table 1 shows the summary statistics of annual percentage yields at bimonthly frequency of all the twenty-three commodities listed in Appendix Table B1. The average (mean) values of percentage yield are positive for the majority of commodities. (The exceptions are

16 In order to address concerns that the results might be sensitive to the method of constructing a commodity index, we also do the analysis for another price-weighted index: a pure (non-scaled) price-weighted index. We do not report descriptive statistics for this index, but the regression results are shown in the Supplementary material. We find similar results to those discussed in Section 4.1.

17 The annual return is compounded from bimonthly return. The annual net convenience yield is calculated as $D_t = R_t * \tilde{F}_{t-6} - \tilde{F}_t$, where D_t refers to annual net convenience yield and R_t refers to annual return. Using annualized returns and yield growth implies that we have "overlapping" data, as in, for example, Cochrane (2011), for which the statistics in the analysis are corrected. This approach is common because there can be seasonality in dividend yields (and certainly in convenience yields). To mitigate this concern, yields are therefore commonly accumulated for a year. We have conducted analyses at higher aggregation frequency though, and find qualitatively similar results.

18 It is worth noting that for any choice of F_0 the resulting neglog dividend yield variable will be *identical*. The choice, however, does affect the constant of linearization, ρ, and thus the accuracy of the linear approximation of the return decomposition.

19 Accordingly, we estimate $D_0 = 8.651$ for the price-weighted index and $D_0 = 4.807$ for the equally weighted index.

Table 1. Summary statistics of annual percentage net convenience yield.

This table shows the annual percentage net convenience yields of individual commodities at bimonthly frequency. Commodities are ordered by average percentage net convenience yield. We roll over returns in 1 year and calculate the gross percentage yield $Y_t = D_t/\tilde{F}_t = R_t\tilde{F}_{t-1}/\tilde{F}_t - 1$, where $\tilde{F}_t = F_t/(1+rf_t)$, where F_t is the futures price at time t of contract maturing at time $t+1$ and rf_t is the risk-free interest rate. Here one period refers to 1 year. The data spans of individual commodities are shown in Appendix Table B1.

Commodity	Obs	Mean (%)	Std. Dev. (%)	Min (%)	Max (%)
Gasoline	229	10.5	15.75	−21.18	55.56
Copper	381	8.1	14.69	−6.37	64.59
Soybean meal	375	7.53	13.58	−16.24	72.35
Lean hogs	326	5.99	19.54	−32.03	50.89
Live cattle	349	5.36	9.99	−14.15	33.27
Crude oil	239	5.21	16.33	−30.15	46.96
Heating oil	265	5.12	14.42	−23.49	40.95
Soybean	382	5.08	11.26	−8.8	73.7
Soybean oil	382	5.03	16.76	−9.49	108.16
Gas oil	220	4.77	14.09	−21.15	45.34
Cotton	375	4.17	17.93	−22.64	105.69
Orange juice	333	3.56	15.49	−22.45	64.94
Feeder cattle	277	3.11	8.84	−16.47	22.75
Coffee	302	1.9	19.24	−21.5	78.58
Cocoa	380	1.76	16.7	−17.71	79.31
Oats	381	−0.37	16.26	−22.71	80.27
Gold	289	−0.54	1.01	−4.57	1.59
Silver	334	−1.3	1.54	−6.01	3.05
Lumber	319	−1.43	22.96	−39.12	96.22
Corn	382	−1.46	11.42	−19.67	41.78
Wheat	381	−1.71	12.45	−23.22	46.07
Rough rice	203	−6.29	11.38	−26.88	33.6
Natural gas	197	−12.45	22.96	−55.13	66.88

corn, wheat, oats, rough rice, lumber, silver, gold, and natural gas.) So, for most commodities, the average convenience yield is larger than its storage cost. Gasoline has the highest average percentage yield (10.50 percent), followed by copper and soybean meal. The percentage yield of natural gas is the most volatile energy commodity with a standard deviation of 22.96 percent, in line with the findings of Brooks, Prokopczuk, and Wu (2013). Standard deviations are generally large relative to the mean, and all commodities exhibit periods during which the net percentage yield is negative.

As discussed in Section 3.1, we construct spot and futures price indices with the twenty-three commodities.[20] The variation of annual percentage yield, neglog annual percentage yield, log annual return, and neglog annual yield growth of the price-weighted index are shown in figure 1 and for the equally weighted index in figure 2. In general, the variation of the annual percentage yield is quite similar to that of the neglog annual percentage yield defined in this article, which supports the use of the neglog annual percentage yield. When the market is in contango ($Y_t < rf_t$),[21] commodities tend to have lower percentage yields as well as returns, as indicated by shaded areas in figures 1 and 2.

Table 2 shows the summary statistics of the annual percentage yields, annual capital gains, annual log commodity returns, annual neglog yield growth, and annual neglog percentage yield for the price-weighted and equally weighted indices. Interestingly, the

[20] The weights of individual commodities in the price-weighted index are shown in the Supplementary material.

[21] The sign used to determine the market condition is the basis. Since we use annualized variables in this article, we use annualized basis rather than bimonthly basis. We infer the annual basis from the annual percentage yield. The percentage yield is defined as $Y_t = D_t/\hat{F}_t = S_t(1+rf_t)/F_t - 1$ which approximates to $\ln[S_t/F_t] + rf_t$. Therefore, we define $Y_t < rf_t$ as the market being in contango.

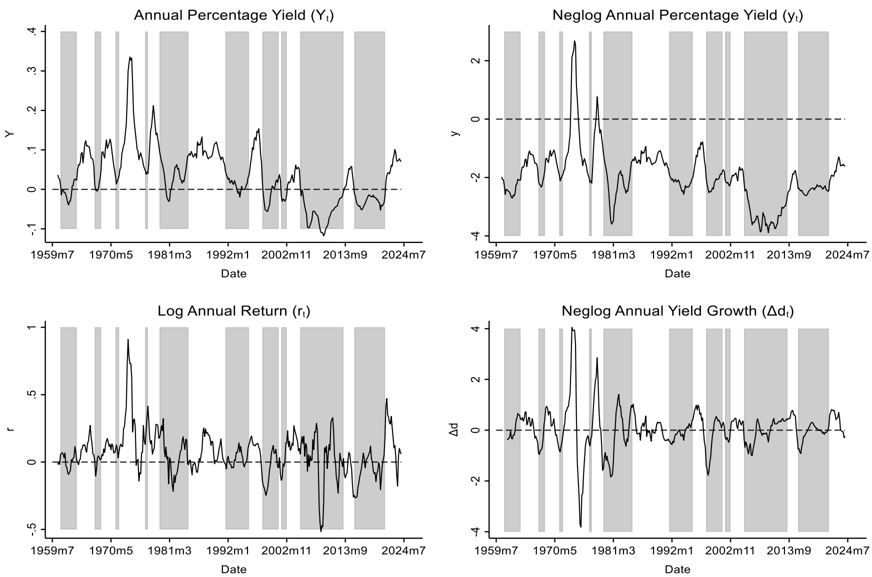

Figure 1. The variation of key variables for the price-weighted commodity index.

The variables are at bimonthly frequency. The annual percentage yield is calculated as $Y_t = \frac{D_t}{\tilde{F}_t}$, where D_t is the level of the annual net convenience yield and $\tilde{F}_t = \frac{F_t}{1+rf_t}$. The neglog annual percentage yield is calculated as $y_t = \text{nl}_\alpha(\hat{D}_t) - \ln(\hat{F}_t) + y_0$ and the neglog yield growth as $\Delta d_{t+1} = \text{nl}_\alpha(\hat{D}_{t+1}) - \text{nl}_\alpha(\hat{D}_t) + \mu$, where $\text{nl}_\alpha(\hat{D}_t) = \alpha \text{sgn}(\hat{D}_t)\ln(\frac{|\hat{D}_t|}{\alpha}+1)$, $\hat{D}_t = \frac{D_t}{D_0 e^{\mu t}}$, $\hat{F}_t = \frac{\tilde{F}_t}{F_0 e^{\mu t}}$, and $y_0 = \ln(D_0/F_0)$. The log annual return is $r_t = \ln(R_t)$, where R_t is gross annual return. The shaded area refers to periods when $Y_t < rf_t$, that is, the market is in contango. The annual percentage yield is from July 1960 to January 2024, and the annual return is from July 1960 to January 2024. The annual yield growth is from July 1961 to January 2024.

first-order autocorrelation of the neglog percentage yields is 0.617 for the price-weighted index and 0.669 for the equally weighted index, so these are less persistent than the dividend yield of common stock (its autocorrelation is usually larger than 0.9; see also the Supplementary material).[22]

4. Results

4.1 The predictability of commodity returns and yield growth

Using the generalized Campbell–Shiller decomposition, we relate the volatility of the percentage yield to the predictability of future returns, future yield growth, and future percentage yields. If neither of these three components were predictable, the percentage yield should be constant, which is obviously not as shown in figures 1 and 2 in Section 3.3. Therefore, the percentage yield *must* forecast either future commodity returns, future yield growth, future percentage yield, or some combination.

The regression results of equation (18) for the price-weighted and equally weighted indices are shown in Table 3. Here one period refers to a year. Since we use overlapping data at bimonthly frequency, the corresponding *t*-statistics of the coefficients are corrected with the Newey–West correction (Newey and West 1987), in line with existing studies (e.g., Kaniel, Saar, and Titman 2008; Maio and Santa-Clara 2015; Cieslak and Povala 2016;

[22] Appendix Table B3 compares the variance of the neglog yield growth and neglog percentage yield with their corresponding log variables based on the subsample of data points with positive net convenience yields. Again, the *variance* of the neglog variables is similar to their corresponding log variables, which is reassuring in light of the subsequent regression analyses.

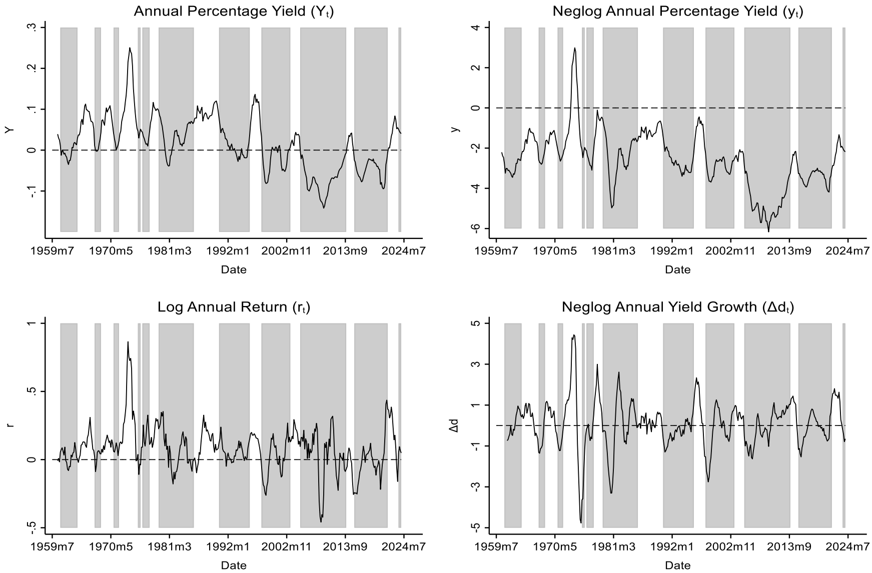

Figure 2. The variation of key variables for the equally weighted commodity index.

The variables are at bimonthly frequency. The annual percentage yield is calculated as $Y_t = \frac{D_t}{\tilde{F}_t}$, where D_t is the level of the annual net convenience yield and $\tilde{F}_t = \frac{F_t}{1+rf_t}$. The neglog annual percentage yield is calculated as $y_t = \mathrm{nl}_\alpha(\hat{D}_t) - \ln(\hat{F}_t) + y_0$ and the neglog yield growth as $\Delta d_{t+1} = \mathrm{nl}_\alpha(\hat{D}_{t+1}) - \mathrm{nl}_\alpha(\hat{D}_t) + \mu$, where $\mathrm{nl}_\alpha(\hat{D}_t) = \alpha \mathrm{sgn}(\hat{D}_t)\ln\left(\frac{|\hat{D}_t|}{\alpha} + 1\right)$, $\hat{D}_t = \frac{D_t}{D_0 e^{\mu t}}$, $\hat{F}_t = \frac{\tilde{F}_t}{F_0 e^{\mu t}}$, and $y_0 = \ln(D_0/F_0)$. The log annual return is $r_t = \ln(R_t)$, where R_t is gross annual return. The shaded area refers to periods when $Y_t < rf_t$, that is, the market is in contango. The annual percentage yield is from July 1960 to January 2024, and the annual return is from July 1960 to January 2024. The annual yield growth is from July 1961 to January 2024.

Table 2. Summary statistics of annual percentage yield, capital gains, commodity return, and yield growth.

This table displays the summary statistics of annual percentage yield ($Y_t = D_t/\tilde{F}_t$), annual capital gain ($\Delta F = \tilde{F}_t/\tilde{F}_{t-1} - 1$), annual commodity return ($r_t = \ln(R_t)$), annual neglog yield growth ($\Delta d_{t+1} = \mathrm{nl}_\alpha(\hat{D}_{t+1}) - \mathrm{nl}_\alpha(\hat{D}_t) + \mu$), and neglog annual percentage yield ($y_t = \mathrm{nl}_\alpha(\hat{D}_t) - \ln(\hat{F}_t) + y_0$), $\mathrm{nl}_\alpha(\hat{D}_t) = \alpha \mathrm{sgn}(\hat{D}_t)\ln(\frac{|\hat{D}_t|}{\alpha} + 1)$, $\hat{D}_t = \frac{D_t}{D_0 e^{\mu t}}$, $\hat{F}_t = \frac{\tilde{F}_t}{F_0 e^{\mu t}}$, and $y_0 = \ln(D_0/F_0)$. The variables are at bimonthly frequency. We set $\alpha = 100$, for the price-weighted index the estimated values for the constants are $\mu = 0.005$, $F_0 = 100.5$, and $D_0 = 8.651$, and for the equally weighted index$\mu = 0.011$, $F_0 = 95.6$, and $D_0 = 4.807$. The symbol ϕ represents the first-order autocorrelation.

Variable	Obs	Mean	Std. Dev.	Min	Max	ϕ
Panel A: Price-weighted index						
Y	382	0.036	0.073	−0.118	0.335	0.650
ΔF	382	0.042	0.154	−0.331	0.913	0.011
r	382	0.064	0.166	−0.517	0.910	0.150
Δd	376	0.041	0.859	−3.820	4.053	−0.065
y	382	−1.943	0.944	−3.892	2.679	0.617
Panel B: Equally weighted index						
Y	382	0.018	0.067	−0.142	0.251	0.662
ΔF	382	0.078	0.159	−0.278	0.953	0.073
r	382	0.081	0.163	−0.459	0.865	0.165
Δd	376	0.076	1.183	−4.769	4.438	−0.044
y	382	−2.563	1.406	−6.167	2.988	0.669

Table 3. Regression results of the vector autoregressions.

This table shows the regression results, h refers to a h-year horizon for the regressions, so $h=1$ is a 12-month return, for example. y refers to the (neglog) percentage yield. For commodity returns, the coefficients are the estimates of $\sum_{j=1}^{h} r_{t+j}\rho^{j-1} = a_r^{(h)} + \beta_r^{(h)} y_t + \varepsilon_{t+h}^r$. For the yield growth, the coefficients are the estimates of $\sum_{j=1}^{h} \Delta d_{t+j}\rho^{j-1} = a_d^{(h)} + \beta_d^{(h)} y_t + \varepsilon_{t+h}^d$. For the bubble regression, the coefficients are the estimates of $y_{t+h} = a_y^{(h)} + \beta_y^{(h)} y_t + \varepsilon_{t+h}^y$. When the horizon is ∞, the coefficients are inferred from the coefficients of the 1-year horizon; $\beta_r^{lr} = \frac{\beta_r}{1-\rho\beta_y}$, $\beta_d^{lr} = \frac{\beta_d}{1-\rho\beta_y}$, and $\beta_y^{lr} = \lim_{h\to\infty}\beta_y^h = 0$. The parameter $\rho = \frac{1}{1+e^{y_0}}$, and we have $\rho = 0.921$ for the price-weighted index and $\rho = 0.952$ for the equally weighted index. The t-statistics are in parentheses and are corrected with the Newey–West correction. *, **, and *** denote statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

	Return		Yield growth		Bubble				
h	$\beta_r^{(h)}$	R_r^2	$\beta_d^{(h)}$	R_d^2	$\beta_y^{(h)}$	R_y^2	$\rho^h\beta_y^{(h)}$	$\beta_r^{(h)} - \beta_d^{(h)} + \rho^h\beta_y^{(h)}$	N
Panel A: Price-weighted index									
1	0.049**	0.079	−0.378***	0.174	0.617***	0.381	0.568	1.00	376
	(2.37)		(−3.28)		(5.67)				
2	0.069**	0.073	−0.677***	0.323	0.294**	0.087	0.250	1.00	370
	(2.57)		(−5.28)		(2.25)				
3	0.101***	0.115	−0.710***	0.363	0.236	0.056	0.185	1.00	364
	(3.73)		(−5.27)		(1.62)				
4	0.132***	0.152	−0.570***	0.321	0.405***	0.165	0.291	0.99	358
	(3.68)		(−5.11)		(3.42)				
5	0.165***	0.189	−0.632***	0.370	0.296***	0.088	0.196	0.99	352
	(3.92)		(−5.77)		(2.71)				
∞	0.114		−0.875		0		0	0.99	
Panel B: Equally weighted index									
1	0.030**	0.067	−0.331***	0.157	0.669***	0.446	0.637	1.00	376
	(2.17)		(−3.16)		(6.56)				
2	0.044**	0.064	−0.620***	0.301	0.367**	0.134	0.333	1.00	370
	(2.36)		(−4.38)		(2.53)				
3	0.067***	0.108	−0.651***	0.332	0.323**	0.104	0.279	1.00	364
	(3.24)		(−4.54)		(2.14)				
4	0.096***	0.168	−0.564***	0.306	0.408***	0.164	0.335	1.00	358
	(3.64)		(−4.91)		(3.41)				
5	0.122***	0.215	−0.612***	0.348	0.332***	0.108	0.260	0.99	352
	(3.88)		(−6.30)		(3.49)				
∞	0.082		−0.911		0		0	0.99	

Golez and Koudijs 2018). The scatter plots associated with the return, yield growth, and bubble regressions for the 1-year horizon are shown in figure 3. The regression scatter plots for longer horizons are shown in the Supplementary material. The long-horizon commodity returns and yield growth are weighted by ρ^{j-1}, where $1 \le j \le h$, h refers to the regression horizon, $\rho = 0.921$ for the price-weighted index and $\rho = 0.952$ for the equally weighted index in this case.

As shown in Table 3, for both indices, the return coefficients β_r are positive (which can also be seen from figure a in panel A and panel B in fig. 3) and the yield growth coefficients β_d are negative[23] (which can also be seen from figure b in panel A and panel B in fig. 3),

[23] OLS estimates for the coefficients in these types of regressions are consistent, but can be subject to bias. It is well known that OLS estimates of an AR(1) process are consistent, but not unbiased; moreover, the bias in the coefficients of the return and yield growth regressions is proportional to this bias if the error terms are cross-correlated (see, e.g., Stambaugh 1999). Unbiased estimates for the price-weighted index for the 1-year horizon are $\beta_r = 0.059, \beta_d = -0.348, \beta_y = 0.639$. These estimates are based on non-overlapping annual data (T=63), using the formulas $\beta_i^{adj} = \hat{\beta}_i - \gamma_i(\beta_y^{adj} - \hat{\beta}_y), \gamma_i = cov(\varepsilon_i, \varepsilon_y)/var(\varepsilon_y), i = r, d$ and $\beta_y^{adj} = \hat{\beta}_y + (1+\hat{\beta}_y)/T$, where the

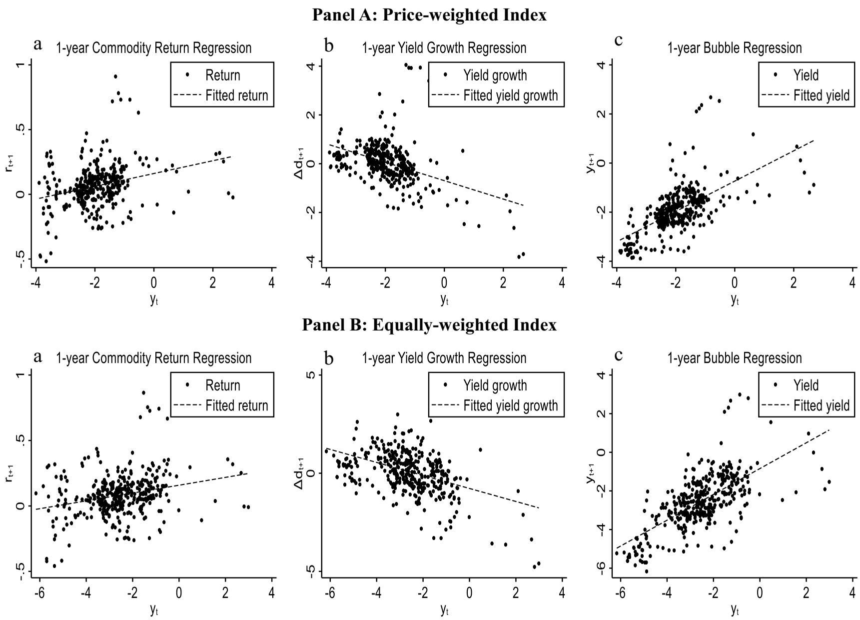

Figure 3. One-year horizon scatter plots for the price-weighted and equally weighted indices.

For commodity returns, the scatter plots relate to the estimation of $r_{t+1} = a_r + \beta_r y_t + \varepsilon^r_{t+1}$. For the yield growth, the scatter plots relate to the estimation of $\Delta d_{t+1} = a_d + \beta_d y_t + \varepsilon^d_{t+1}$. For the bubble, the scatter plots relate to the estimation of $y_{t+1} = a_y + \beta_y y_t + \varepsilon^y_{t+1}$.

which is consistent the implied relationship from equation (13) and the standard economic interpretation; the percentage yield is positively related to expected returns and negatively related to expected yield growth. The long-run coefficients (i.e., when h goes to infinity) in Table 3 are inferred from the one-year coefficients, as is standard in the predictability literature (e.g., Engsted and Pedersen 2010). The sum of the coefficients reported in the one to last column is always very close to 1, implying that the approximate identity (15) is accurate. In the long-run, the predictability of the percentage yield disappears and percentage yield volatility is driven entirely by the predictability of both returns and yield growth. These results show that the contribution of expected yield growth (i.e., payoffs) is much larger than the contribution of expected returns (i.e., discount rates). A high current commodity price reflects a high expected future payoff as well as a low discount rate.

In particular, Table 3 shows that the return coefficients β_r increase slowly with horizon. The R^2_r for the return regressions increase slowly as well, so we do observe the common predictability pattern in commodity returns. Such predictability is supported by many studies that try to forecast commodity returns with other variables, such as financial and macroeconomic variables (e.g., inflation, interest rates, and dividend–price ratio; Chen, Rogoff, and Rossi 2010; Gargano and Timmermann 2014) and commodity-specific factors (e.g., open interest and hedging demand; Hong and Yogo 2012; Acharya, Lochstoer, and Ramadorai 2013). A high current commodity price relates to a low percentage yield, which forecasts a low future return. Therefore, the current price is high because investors expect a low future return (discount rate), which is consistent with the common story for stock prices. However, the contribution of future returns in driving the variation of commodity

superscript *adj* indicates the estimate adjusted for bias. Note that these bias adjustments ignore the theoretical reprices in the long-run is only 11.4 percent for the price-weighted index and 8.2 percent for the equally weighted index.

The absolute values of yield growth coefficients β_d also increase with horizon as shown in Table 3. A high commodity price relates to a low percentage yield, which predicts high yield growth and thus a high net convenience yield. The predictability of yield growth contributes 87.5 percent to the variation of commodity prices in the long-run for the price-weighted index and 91.1 percent for the equally weighted index, which is much larger than the contribution of commodity returns. Therefore, in general, a commodity price is high because investors expect a high future net convenience yield. This stronger predictability of yield growth in commodity markets is different from the predictability patterns observed in stock markets where dividend growth is unpredictable by the dividend yield only. This stylized fact for stock markets has actually been labeled as "discomforting" (e.g., Cochrane 2008; Engsted and Pedersen 2010). Cochrane (2008) mentions that it would be nice if a high stock price reflects the expectation of high future dividend growth.

With regard to commodities, this is the case: our results imply that an investor would like to hold a commodity if he expects a large net convenience yield (payoff) in the future. In particular, if the current net convenience yield is negative and the current price is high, investors expect a higher net convenience yield in the future, since investors would not buy and hold a commodity with a structural negative net convenience yield at a positive price.[24] Therefore, in light of potential negative yields, the predictability of commodity yield growth is perhaps not that surprising, and this is corroborated by the results.

As regards the bubble regressions shown in Table 3, the bubble coefficients β_y are positive. A low percentage yield, implying a high commodity price, predicts a high future commodity price. Investors might hold a commodity not for the net convenience yield, but with the expectation that someone else is willing to pay more for this commodity in the future. But such a "rational bubble" does not exist in the long run, because $\beta_y < 1$ and $\rho\beta_y < 1$. This finding is in line with the studies about (rational) price bubbles on commodity markets, which suggest that such bubbles are short-lived and probably only last for several months (e.g., Figuerola-Ferretti, Gilbert, and McCrorie 2015; Araujo Bonjean and Simonet 2016; Pan 2018; Sharma and Escobari 2018).

4.2 Joint tests for return predictability and yield growth predictability

The analysis so far is based on the VAR regression (18). The commodity return, yield growth, and bubble regressions are treated as separate time series. For example, for the regression of commodity returns, the null hypothesis we test is $H_0 : \beta_r = 0$, while β_d and β_y could be anything. However, we know that β_y cannot be too large in a coherent world; otherwise, the percentage yield will explode. As argued by Cochrane (2008), when we study price variation with the decomposition, we should not ask "Are returns predictable?" or "Is yield growth predictable?" Instead, we should ask "which of the return and yield growth is predictable?" According to equation (15), a null hypothesis that the return is not predictable must imply that the yield growth is predictable and vice versa.

In order to jointly test the predictability of commodity returns and yield growth, we conduct Monte Carlo simulations, as discussed in Section 2.4, to study the joint distribution of the coefficients β_r and β_d. We consider two null hypotheses. The first one is that the return is unpredictable and yield growth is predictable. The second one is that the return is predictable and the yield growth is unpredictable. We conduct 50,000 simulations for each null hypothesis. We first simulate percentage yields with the sample estimate β_y (one-period estimation with full sample shown in Table 3) and the correlation and covariance matrix of

24 According to the present-value framework, it is not possible that the net convenience yield (payoff) is expected to be permanently negative when the current price is positive, as the price should equal the expected discounted future payoff.

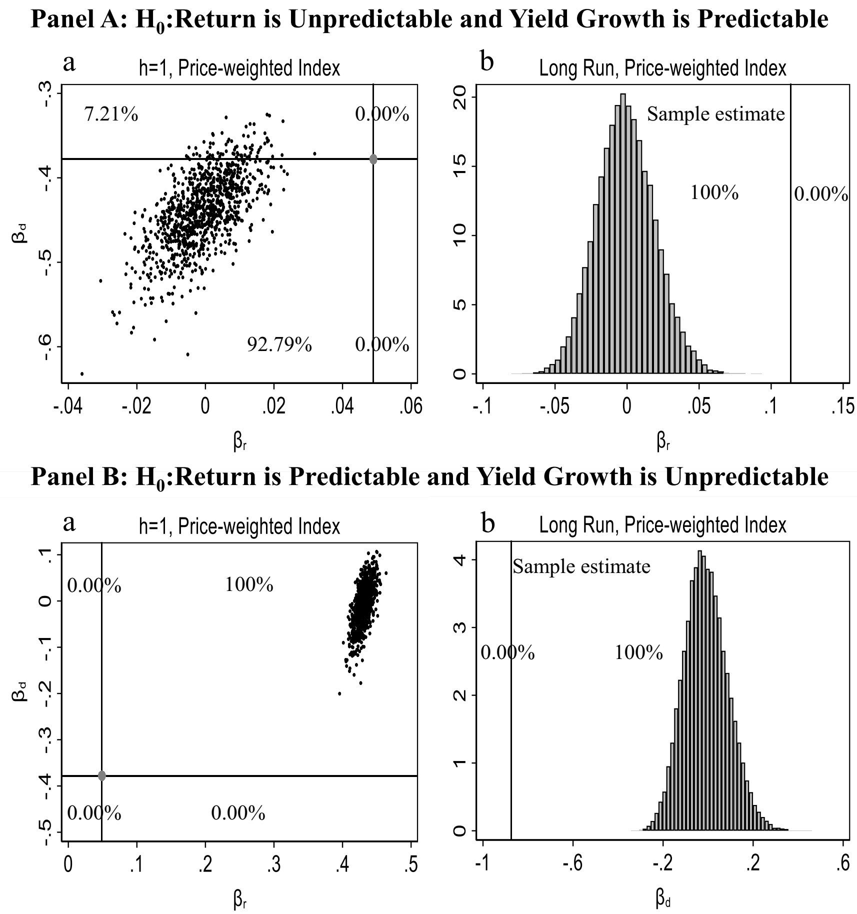

Figure 4. Joint distribution of 1-year horizon and long-run return and yield growth coefficients for the price-weighted index.

h refers to regression horizons. The lines and the large gray dots are associated with the sample estimates of 1-year horizon coefficients. One thousand simulation results are selected randomly and are displayed for clarity in the joint distribution plots. The percentages are the fractions of 50,000 simulation points that fall in each quadrant. For the distribution of long-run coefficients, 50,000 simulation points are used. The vertical lines in the histograms give the sample estimate of long-run coefficients.

ε^d and ε^y shown in Appendix Table B4. In each Monte Carlo simulation, we estimate equation (18) and collect corresponding coefficients.

The joint distributions of the return coefficient β_r and yield growth coefficient β_d for the 1-year horizon and the distributions of long-run coefficients β_r^{lr} and β_d^{lr} for the price-weighted and equally weighted indices are plotted in figures 4 and 5. Panel A in figures 4 and 5 displays the simulation results for the first null hypothesis that the return is unpredictable and yield growth is predictable. In this case, none of the 50,000 simulations produces a return coefficient that is larger than the sample estimate (i.e., to the right of the vertical line in scatter plot *a* of Panel A in figs. 4 and 5). All the simulated points are centered in the area with negative yield growth coefficients, in line with the sign of the sample value. The first null hypothesis implies that all the variation of the percentage yield comes from time-varying future yield growth in the long-run, that is, $\beta_r^{lr} = 0$ and $-\beta_d^{lr} = 1$. Under

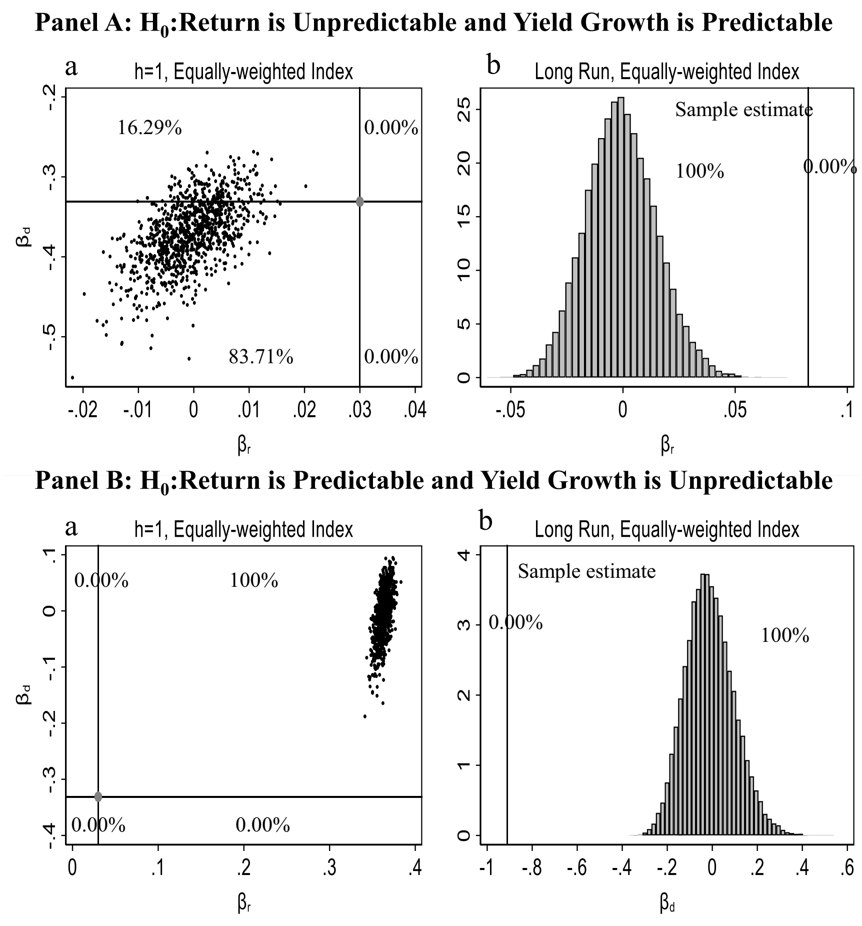

Figure 5. Joint distribution of 1-year horizon and long-run return and yield growth coefficients for the equally weighted index.

h refers to regression horizons. The lines and the large gray dots are associated with the sample estimates of 1-year horizon coefficients. One thousand simulation results are selected randomly and are displayed for clarity in the joint distribution plots. The percentages are the fractions of 50,000 simulation points that fall in each quadrant. For the distribution of long-run coefficients, 50,000 simulation points are used. The vertical lines in the histograms give the sample estimate of long-run coefficients.

this null, again no simulation produces a long-run return coefficient larger than the sample estimate, which can be seen from histogram *b* in Panel A in figures 4 and 5. It is not possible to obtain a return coefficient larger than the sample estimate by pure chance if the return is truly unpredictable. Therefore, the null hypothesis that returns are unpredictable is rejected for the price-weighted and equally weighted indices.

The same logic applies to the second null hypothesis that the return is predictable and yield growth is unpredictable, which can be seen in Panel B in figures 4 and 5. As shown in scatter *a* in Panel B of figures 4 and 5, all points are centered above the horizontal line. The yield growth coefficients should center around zero if it is not forecastable. All the simulated yield growth coefficients are much larger than the sample estimate (recall that the sample estimate of yield growth coefficient is negative). For the 1-year horizon regressions (scatter *a* in Panel B in figures 4 and 5), none of the simulated yield growth coefficients is smaller than the sample estimate. All the simulated return coefficients are positive for the case when the return is predictable, in line with the sign of the sample estimate. The second null hypothesis suggests that all variation in the percentage yield comes from time-varying expected returns rather than the yield growth in the long-run, that is, $\beta_r^{lr} = 1$ and $\beta_d^{lr} = 0$. With this null as shown in histogram *b* in Panel B of figures 4 and 5, the simulated long-run yield growth coefficient cannot be smaller than the sample estimate by pure chance if the yield growth is truly not predictable. So the yield growth must be predictable if we have such a small (more negative) coefficient for yield growth. Therefore, the null hypothesis that yield growth is not predictable is rejected for the price-weighted and equally weighted indices. The joint distribution of return and yield growth coefficients of longer-horizon regressions is shown in the Supplementary material. These results also suggest that both return and yield growth are predictable because no simulation generates a larger return coefficient than the sample estimate and a smaller yield growth coefficient than the sample estimate.

4.3 What drives the predictability?—A technical explanation

We find that the variation of the percentage yield comes from both expected future yield growth and expected future returns and that expected returns only contribute slightly in the long run, which is different from the stylized facts for stock markets and most other financial markets. In this section, we provide a technical answer to what drives the predictability of commodity returns and yield growth.

The joint distribution of β_y and β_r and the joint distribution of β_y and β_d under the two null hypotheses for the price-weighted and equally weighted indices is displayed in figures 6 and 7. Figures 6 and 7 show only a limited correlation between the return coefficient β_r and the bubble coefficient β_y (see scatter *a* in Panel A and Panel B in figs 6 and 7), but a strong and positive correlation between the yield growth coefficient β_d and the bubble coefficient β_y (see scatter *b* in Panel A and Panel B in figs 6 and 7). This is mainly because the yield growth shocks and bubble (percentage yield) shocks are strongly and positively correlated, which can be seen from the error covariance and correlation matrix shown in Appendix Table B4. Therefore, a shock that produces a small bubble coefficient corresponds to a shock that produces a small yield growth coefficient, which implies high predictability of yield growth (recall that the yield growth coefficient is negative). With the restriction of identity (15), a strong correlation between β_y and β_d implies that the β_y and β_r are not correlated or, if at all, are only slightly correlated, which is evidenced by figures 6 and 7. The strong positive correlation between yield growth and bubble shocks explains the higher predictability of yield growth compared to commodity returns.

5. Robustness checks

We conduct two robustness checks. First, to verify that neglog-linearization is not driving the results, we consider longer time horizons in Section 5.1 where net convenience yields are strictly positive (it turns out that with our sample this requires considering 4-year returns). While we lose statistical power at this lower frequency, it enables us to do the standard Campbell–Shiller decomposition and compare it qualitatively to the neglog results. Second, commodities vary in terms of how much they are subject to seasonality, so in Section 5.2, we investigate whether seasonal commodities drive the predictability results.

5.1 Classic Campbell–Shiller decomposition with strictly positive yields by using a subsample of commodities and 4-year returns

In order to solve the technical issue related to the negative net convenience yield while retaining the economic properties necessary for the analysis, we redefined the log percentage yield and log dividend growth using the neglog transform. However, the redefined

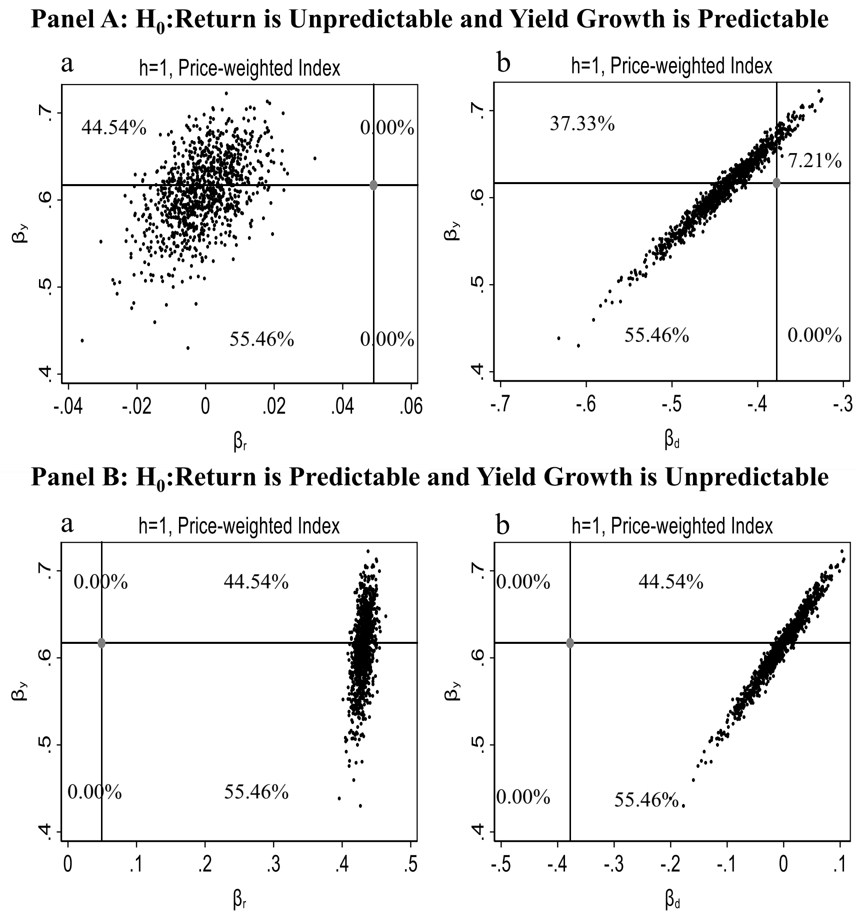

Figure 6. Joint distribution of 1-year return, yield growth, and bubble coefficients for the price-weighted index.

h refers to regression horizons. The lines and the large gray dots are associated with the sample estimates of 1-year horizon coefficients. One thousand simulation results are selected randomly and are displayed for clarity in the joint distribution plots. The percentages are the fractions of 50,000 simulation points that fall in each quadrant.

neglog percentage yield and neglog dividend growth are somewhat different from those commonly constructed with logs for positive yields. In order to address some of the concerns arising from the differences in these definitions, we conduct the classic Campbell–Shiller decomposition based on a subsample of commodity prices with longer holding-period returns such that we obtain strictly positive net convenience yields.[25] Using this sample, we check for robustness by comparing the results based on the standard

[25] To come to this index, we conduct a "grid search," where the trade-off is between frequency and number of commodities in the index. We start with the highest frequency possible, adding and removing commodities until we obtain an index with strictly positive yields accumulated over the chosen frequency. For higher frequencies, this results in only a few commodities in the index. Since theoretically there is not a statistical difference between long and short horizon regressions, we feel that we should focus more on having enough (at least 20) commodities in the index to allow for fair comparison with the baseline results.

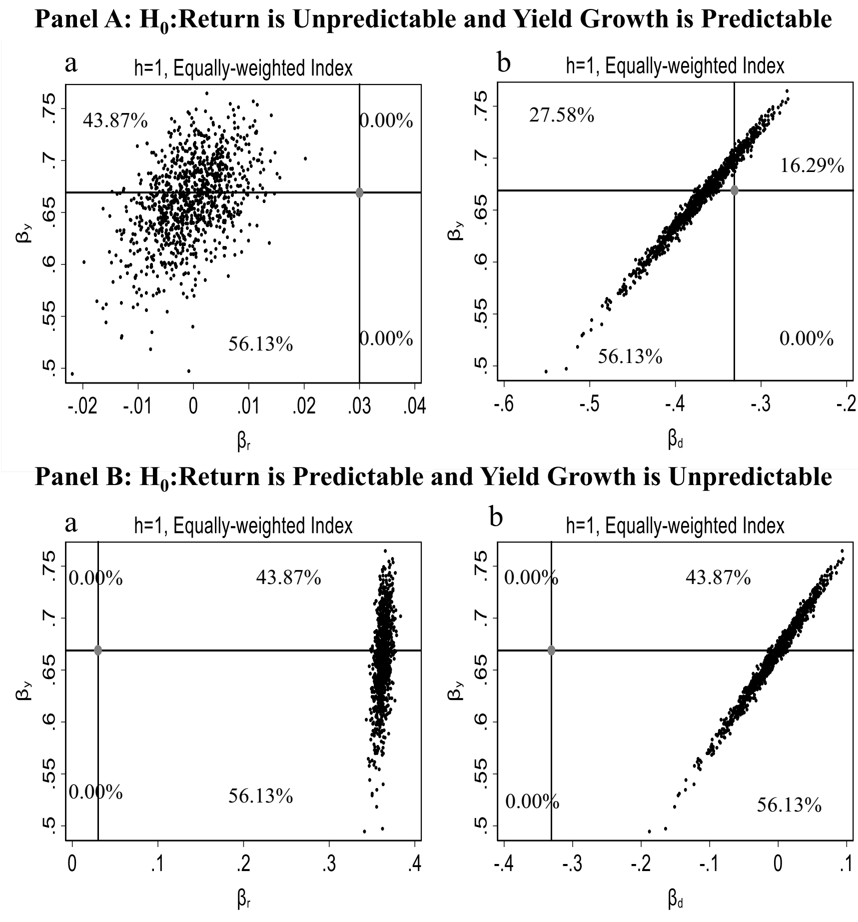

Figure 7. Joint distribution of 1-year return, yield growth, and bubble coefficients for the equally weighted index.

h refers to regression horizons. The lines and the large gray dots are associated with the sample estimates of 1-year horizon coefficients. One thousand simulation results are selected randomly and are displayed for clarity in the joint distribution plots. The percentages are the fractions of 50,000 simulation points that fall in each quadrant.

Campbell–Shiller approach with the results based on the generalized Campbell–Shiller decomposition discussed in Section 2.2.[26]

As it turns out, to obtain a somewhat representative sample that only exhibits strictly positive net convenience yields requires working with the gross net convenience yields for a period of 4 years and a buy-and-hold index that includes seventeen commodities. The yields remain positive for the period between July 1963 and March 2002. These seventeen commodities include soybean oil, soybean meal, rough rice, coffee, orange juice, live cattle, feeder cattle, lean hog, cotton, silver, copper, gold, gasoil, natural gas, crude oil, gasoline, and heating oil. With the classic Campbell–Shiller decomposition, we relate the variation of the standard log percentage yield to the predictability of returns and standard log yield

[26] For the sake of completeness, the derivation of the standard Campbell–Shiller decomposition applied to commodity returns is reported in the supplementary material.

growth. In sum, we have $y_t = r_{t+1} - \Delta d_{t+1} + \rho y_{t+1} - \kappa$, where $y_t = \ln(Y_t)$ and $\Delta d_{t+1} = \ln(\frac{D_{t+1}}{D_t})$.

Using the subsample from July 1963 to March 2002, that is, when the 4-year net convenience yields are always positive, the regression results of the classic and generalized Campbell–Shiller decomposition are shown in Appendix Table B5. In this case, $h = 1$ refers to a 4-year horizon. The corresponding t-statistics of the coefficients are corrected with the Newey–West correction (Newey and West 1987). As shown in Panel A in Appendix Table B5 for the classic Campbell–Shiller decomposition, the return coefficient β_r is positive and the yield growth coefficient β_d is negative. In the long run, the variation in the percentage yield comes from both the predictability of returns and yield growth. These results suggest that around 36 percent of the variation in the percentage yield can be attributed to time-varying expected returns (discount rates), while the other 63.8 percent of the variation comes from time-varying expected net convenience yield growth (payoffs). The contribution of expected payoffs is thus much higher than that of discount rates. Panel B in Appendix Table B5 provides quantitatively similar results for the generalized Campbell–Shiller decomposition. These results are consistent with the results discussed in Section 4, which validates the use of our definition of neglog percentage yield and neglog dividend growth as discussed in Section 2.2.

5.2 Do seasonal commodities drive the results?

In this section, we shed some light on whether seasonality in the net convenience yield matters to the predictability of return and yield growth. Even though we aggregate convenience yields over a year, there might be overlooked aspects of seasonality features in the data that drive the observed predictability patterns. Seasonality is not surprising for the net convenience yield, as, for example, agriculture commodities have harvests and some energy commodities have seasonal demand peaks as well (e.g., Fama and French 1987; Brooks, Prokopczuk, and Wu 2013). It is also a common phenomenon for dividends at monthly or quarterly frequency (e.g., Polimenis and Neokosmidis 2016; Asimakopoulos et al. 2017).

In order to investigate the effect of seasonality, we divide the commodities into a seasonal and nonseasonal group based on studies about seasonality in commodity prices (e.g., Cartea and Williams 2008; Karali and Ramirez 2014; Arismendi et al. 2016; Hevia, Petrella, and Sola 2018). The seasonal commodity set includes heating oil, gasoline, crude oil, natural gas, gasoil, orange juice, rough rice, soybean, soybean meal, soybean oil, corn, wheat, coffee, cocoa, oats, cotton, and lumber. The other commodities belong to the nonseasonal group, including gold, copper, silver, feeder cattle, live cattle, and lean hogs. The upbringing and slaughtering of livestock might happen anytime (e.g., Dimpfl, Flad, and Jung 2017), so we classify livestock into the nonseasonal group.

The regression results for the price-weighted and equally weighted indices for the seasonal group are shown in Table 4. The return coefficient β_r is positive, and the yield growth coefficient β_d is negative, which is consistent with the results shown in Table 3. The volatility in the percentage yield comes from the predictability of expected yield growth and expected returns. Again, in the long run ($h = \infty$), the contribution of expected yield growth is much larger than the contribution of expected returns—the current commodity price is high mainly because investors expect a high future payoff (the net convenience yield) and to a lesser extent it reflects a low discount rate (expected return).

In the nonseasonal group, the regression results for the price-weighted and equally weighted indices are shown in Table 5. Again, the return coefficient β_r is positive and the yield growth coefficient β_d is negative. Also, for this group of commodities, the expected yield growth contributes much more than the future returns in driving the volatility of the percentage yield. These results are in line with the results shown in Tables 3 and 4.

We also conduct 50,000 Monte Carlo simulations to jointly test the predictability of return and yield growth in the seasonal and nonseasonal groups. Here, we only proceed with

Table 4. Regression results of the vector autoregressions for the seasonal group.
This table shows the regression results for the seasonal group. For the price-weighted index, $\rho = 0.895$. For the equally weighted index, $\rho = 0.934$. The letter h refers to a h-year horizon for the regressions. y refers to the (neglog) percentage yield. For commodity returns, the coefficients are the estimates of $\sum_{j=1}^{h} r_{t+j}\rho^{j-1} = a_r^{(h)} + \beta_r^{(h)} y_t + \varepsilon_{t+h}^r$. For the yield growth, the coefficients are the estimates of $\sum_{j=1}^{h} \Delta d_{t+j}\rho^{j-1} = a_d^{(h)} + \beta_d^{(h)} y_t + \varepsilon_{t+h}^d$. For the bubble regression, the coefficients are the estimates of $y_{t+h} = a_y^{(h)} + \beta_y^{(h)} y_t + \varepsilon_{t+h}^y$. When the horizon is ∞, the coefficients are inferred from the coefficients of the 1-year horizon; $\beta_r^{lr} = \frac{\beta_r}{1-\rho\beta_y}$, $\beta_d^{lr} = \frac{\beta_d}{1-\rho\beta_y}$, and $\beta_y^{lr} = \lim_{h\to\infty}\beta_y^h = 0$. We set $\alpha = 100$, for the price-weighted index the estimated values for the constants are $\mu = 0.005$, $F_0 = 101.0$, $D_0 = 11.147$ and $\rho = 0.901$, and for the equally weighted $\mu = 0.011$, $F_0 = 92.8$, $D_0 = 6.127$, and $\rho = 0.938$. The t-statistics are in parentheses and are corrected with the Newey–West correction. *, **, and *** denote statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

	Return		Yield growth		Bubble				
h	$\beta_r^{(h)}$	R_r^2	$\beta_d^{(h)}$	R_d^2	$\beta_y^{(h)}$	R_y^2	$\rho^h\beta_y^{(h)}$	$\beta_r^{(h)} - \beta_d^{(h)} + \rho^h\beta_y^{(h)}$	N
Panel A: Price-weighted index									
1	0.053**	0.063	−0.386***	0.177	0.616***	0.377	0.555	0.99	376
	(2.22)		(−3.25)		(5.53)				
2	0.080***	0.069	−0.704***	0.328	0.260*	0.067	0.211	0.99	370
	(2.64)		(−5.05)		(1.87)				
3	0.118***	0.112	−0.735***	0.368	0.195	0.037	0.142	1.00	364
	(3.91)		(−5.60)		(1.36)				
4	0.145***	0.133	−0.619***	0.343	0.346***	0.118	0.228	0.99	358
	(3.55)		(−5.42)		(2.83)				
5	0.172***	0.155	−0.674***	0.395	0.243**	0.059	0.144	0.99	352
	(3.54)		(−5.76)		(2.13)				
∞	0.119		−0.867		0		0	0.99	
Panel B: Equally weighted index									
1	0.032*	0.051	−0.341***	0.161	0.664***	0.438	0.623	1.00	376
	(1.95)		(−3.15)		(6.38)				
2	0.053**	0.067	−0.651***	0.309	0.329**	0.106	0.289	0.99	370
	(2.41)		(−4.32)		(2.15)				
3	0.084***	0.121	−0.684***	0.343	0.274*	0.074	0.226	0.99	364
	(3.37)		(−4.71)		(1.77)				
4	0.112***	0.162	−0.603***	0.324	0.357***	0.124	0.277	0.99	358
	(3.47)		(−5.02)		(2.88)				
5	0.136***	0.190	−0.633***	0.363	0.303***	0.090	0.220	0.99	352
	(3.48)		(−5.73)		(2.85)				
∞	0.084		−0.905		0		0	0.99	

the price-weighted index since the difference in construction of the indices does not seem to impact the results. We do not report the results of this exercise in the main text, but our findings corroborate that these do not seem to be sensitive to seasonality characteristics of commodities.

6. Conclusion

We investigate the importance of time-varying discount rates for commodity prices. As is well documented, the variation in stock prices comes from time-varying expected stock return rather than expected dividend growth. Similar patterns are observed for most other asset classes. We show that in contrast, commodity prices exhibit a unique behavior, strongly predicting future cash flows rather than discount rates.

Table 5. Regression results of the vector autoregressions for the nonseasonal group.
This table shows the regression results for the nonseasonal group. For the price-weighted index, $\rho = 0.918$. For the equally weighted index, $\rho = 0.927$. The letter h refers to a h-year horizon for the regressions. y refers to the (neglog) percentage yield. For commodity returns, the coefficients are the estimates of $\sum_{j=1}^{h} r_{t+j}\rho^{j-1} = a_r^{(h)} + \beta_r^{(h)} y_t + \varepsilon_{t+h}^{r}$. For the yield growth, the coefficients are the estimates of $\sum_{j=1}^{h} \Delta d_{t+j}\rho^{j-1} = a_d^{(h)} + \beta_d^{(h)} y_t + \varepsilon_{t+h}^{d}$. For the bubble regression, the coefficients are the estimates of $y_{t+h} = a_y^{(h)} + \beta_y^{(h)} y_t + \varepsilon_{t+h}^{y}$. When the horizon is ∞, the coefficients are inferred from the coefficients of the 1-year horizon; $\beta_r^{lr} = \frac{\beta_r}{1-\rho\beta_y}$, $\beta_d^{lr} = \frac{\beta_d}{1-\rho\beta_y}$, and $\beta_y^{lr} = \lim_{h\to\infty} \beta_y^h = 0$. We set $\alpha = 100$, for the price-weighted index the estimated values for the constants are $\mu = 0.007$, $F_0 = 112.3$, $D_0 = 10.158$, and $\rho = 0.917$, and for the equally weighted $\mu = 0.010$, $F_0 = 107.6$, $D_0 = 8.576$, and $\rho = 0.926$.The t-statistics are in parentheses and are corrected with the Newey–West correction. *, **, and *** denote statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

	Return		Yield growth		Bubble				
h	$\beta_r^{(h)}$	R_r^2	$\beta_d^{(h)}$	R_d^2	$\beta_y^{(h)}$	R_y^2	$\rho^h\beta_y^{(h)}$	$\beta_r^{(h)} - \beta_d^{(h)} + \rho^h\beta_y^{(h)}$	N
Panel A: Price-weighted index									
1	0.050***	0.071	−0.568***	0.263	0.417***	0.174	0.383	1.00	375
	(3.12)		(−7.16)		(5.42)				
2	0.049**	0.032	−0.789***	0.392	0.196*	0.038	0.165	1.00	369
	(2.08)		(−7.92)		(1.84)				
3	0.072**	0.054	−0.719***	0.393	0.275***	0.075	0.212	1.00	363
	(2.41)		(−9.14)		(3.10)				
4	0.115***	0.116	−0.662***	0.393	0.320***	0.100	0.226	1.00	357
	(2.96)		(−9.43)		(4.68)				
5	0.141***	0.148	−0.728***	0.507	0.208**	0.060	0.135	1.00	351
	(3.02)		(−10.05)		(2.41)				
∞	0.081		−0.920					1.00	
Panel B: Equally weighted index									
1	0.047***	0.079	−0.512***	0.241	0.476***	0.228	0.441	1.00	375
	(3.45)		(−5.92)		(5.48)				
2	0.041**	0.027	−0.763***	0.382	0.228**	0.052	0.196	1.00	369
	(1.97)		(−7.30)		(2.03)				
3	0.059**	0.042	−0.722***	0.394	0.276***	0.075	0.219	1.00	363
	(2.17)		(−9.74)		(3.34)				
4	0.096***	0.096	−0.661***	0.393	0.330***	0.106	0.243	1.00	357
	(2.64)		(−7.99)		(3.86)				
5	0.122***	0.132	−0.720***	0.532	0.233***	0.081	0.159	1.00	351
	(2.70)		(−11.82)		(3.71)				
∞	0.085		−0.915		0		0	1.00	

On commodity markets, the net convenience yield, defined as the benefit accrued to investors by holding the commodity net of storage cost, is the analogue to the dividend on a stock. Moreover, a simple trading strategy can be implemented to collect the net convenience yield as a monetary cash flow. Present-value logic dictates that a commodity's price should equal the net present value of expected net convenience yields, as argued by Pindyck (1993).

With such an equity-like feature, we propose to apply a linearized present-value model to explain the variation in commodity prices. Net convenience yields are frequently negative, which means log-linearization is not suitable. To solve this problem, we introduce alternative measures, the neglog percentage yield, and neglog yield growth. Under these measures, a high price still corresponds to a low percentage yield and a high yield growth still corresponds to a high future net convenience yield, irrespective of whether net convenience yields are positive or negative. We introduce a generalized Campbell–Shiller decomposition using the neglog variables, which allows us to decompose the percentage yield into the expected commodity return, expected yield growth, and the expected future percentage yield (a bubble component). If the percentage yield varies at all, it *must* be because it can predict at least one of the three components.

We find coherent predictability patterns for both commodity returns and yield growth. A high commodity price is attributed to a high expected net convenience yield (i.e., cash flow) as well as a low future return (i.e., discount rate). Notably, the predictability for yield growth is stronger than that for commodity returns. In the long run, commodity price variation seems mainly driven by time-varying expected yield growth. The results do not seem to be sensitive to seasonality characteristics of commodities. In addition, using a subsample of commodities at a lower frequency with only positive net convenience yields, and applying the classic Campbell–Shiller decomposition, corroborates the results with the redefined neglog percentage yield and neglog yield growth.

Again, our findings are different from the stylized facts found for stock markets that only expected stock returns matter for driving variation in stock prices. A technical explanation is that this is in part because the percentage yield is less persistent than the dividend yield and the bubble shocks are strongly related to yield growth shocks on commodity markets. In any case, it is predominantly expected cash flows that determine commodity prices, but it begs the question why commodities behave so differently compared to other assets. We leave it to future research to provide an economic explanation for the predictability patterns in commodity markets.

Acknowledgments

We would like to thank Auke Plantinga, Peter Smid, Bert Scholtens, Wolfgang Bessler, Tommy Lundgren, Robert Lensink, Bjorn Espen Eckbo, participants of the Asset Pricing Meetings, Groningen, and participants of the NHH seminar series, Bergen, for helpful comments.

Supplementary material

Supplementary material is available at *Review of Finance* online.

Funding

This work was partially supported by the China Scholarship Council.

Conflicts of interest: None declared.

Data availability

The proprietary data for the individual commodity futures prices are from the CRB, under a license that prevents us from sharing them publicly. Data for all commodity indices are shared and can be used to reproduce our results (except for the descriptive statistics of individual commodities). Stata codes for generating the results are shared.

References

Acharya, V. V., L. A. Lochstoer, and T. Ramadorai. 2013. "Limits to Arbitrage and Hedging: Evidence from Commodity Markets." *Journal of Financial Economics* 109: 441–65. https://doi.org/10.1016/j.jfineco.2013.03.003.

Araujo Bonjean, C., and C. Simonet. 2016. “Are Grain Markets in Niger Driven by Speculation?” *Oxford Economic Papers* 68: 714–35. https://doi.org/10.1093/oep/gpw012.

Arismendi, J. C., J. Back, M. Prokopczuk, R. Paschke, and M. Rudolf. 2016. “Seasonal Stochastic Volatility: Implications for the Pricing of Commodity Options.” *Journal of Banking & Finance* 66: 53–65. https://doi.org/10.1016/j.jbankfin.2016.02.001.

Asimakopoulos, P., S. Asimakopoulos, N. Kourogenis, and E. Tsiritakis. 2017. “Time-Disaggregated Dividend–Price Ratio and Dividend Growth Predictability in Large Equity Markets.” *Journal of Financial and Quantitative Analysis* 52: 2305–26. https://doi.org/10.1017/S0022109017000643.

Bakas, D., and A. Triantafyllou. 2018. “The Impact of Uncertainty Shocks on the Volatility of Commodity Prices.” *Journal of International Money and Finance* 87: 96–111. https://doi.org/10.1016/j.jimonfin.2018.06.001.

Basak, S., and A. Pavlova. 2016. “A Model of Financialization of Commodities.” *The Journal of Finance* 71: 1511–56. https://doi.org/10.1111/jofi.12408.

Belousova, J., and G. Dorfleitner. 2012. “On the Diversification Benefits of Commodities From the Perspective of Euro Investors.” *Journal of Banking & Finance* 36: 2455–72. https://doi.org/10.1016/j.jbankfin.2012.05.003.

Boons, M., and M. P. Prado. 2019. “Basis-Momentum.” *The Journal of Finance* 74: 239–79. https://doi.org/10.1111/jofi.12738.

Brooks, C., M., Prokopczuk, and Y. Wu. 2013. “Commodity Futures Prices: More Evidence on Forecast Power, Risk Premia, and the Theory of Storage.” *The Quarterly Review of Economics and Finance* 53: 73–85. https://doi.org/10.1016/j.qref.2013.01.003.

Brunetti, C., and D. Reiffen. 2014. “Commodity Index Trading and Hedging Costs.” *Journal of Financial Markets* 21: 153–80. https://doi.org/10.1016/j.finmar.2014.08.001.

Campbell, J. Y., and R. J. Shiller. 1988. “The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors.” *The Review of Financial Studies* 1: 195–228. https://doi.org/10.1093/rfs/1.3.195.

Cartea, A., and T. Williams. 2008. “UK Gas Markets: The Market Price of Risk and Applications to Multiple Interruptible Supply Contracts.” *Energy Economics* 30: 829–46. https://doi.org/10.1016/j.eneco.2007.03.001.

Casassus, J., and P. Collin-Dufresne. 2005. “Stochastic Convenience Yield Implied from Commodity Futures and Interest Rates.” *The Journal of Finance* 60: 2283–331. https://doi.org/10.1111/j.1540-6261.2005.00799.x.

Chen, Y.-C., K. S. Rogoff, and B. Rossi. 2010. “Can Exchange Rates Forecast Commodity Prices?” *The Quarterly Journal of Economics* 125: 1145–94. https://doi.org/10.1162/qjec.2010.125.3.1145.

Christoffersen, P., K. Jacobs, and X. Pan. 2022. “The State Price Density Implied by Crude Oil Futures and Option Prices.” *The Review of Financial Studies* 35: 1064–103. https://doi.org/10.1093/rfs/hhab011.

Cieslak, A., and P. Povala. 2016. “Information in the Term Structure of Yield Curve Volatility.” *The Journal of Finance* 71: 1393–436. https://doi.org/10.1111/jofi.12388.

Cochrane, J. H. 2008. “The Dog That Did Not Bark: A Defense of Return Predictability.” *The Review of Financial Studies* 21: 1533–75. https://doi.org/10.1093/rfs/hhm046.

Cochrane, J. H. 2011. “Presidential Address: Discount Rates.” *The Journal of Finance* 66: 1047–108. https://doi.org/10.1111/j.1540-6261.2011.01671.x.

Dimpfl, T., M. Flad, and R. C. Jung. 2017. “Price Discovery in Agricultural Commodity Markets in the Presence of Futures Speculation.” *Journal of Commodity Markets* 5: 50–62. https://doi.org/10.1016/j.jcomm.2017.01.002.

Domanski, D., and A., Heath. 2007. “Financial Investors and Commodity Markets.” *BIS quarterly review, March*.

Elkamhi, R., and C. Jo. 2023. “Asset Holders' Consumption Risk and Tests of Conditional CCAPM.” *Journal of Financial Economics* 148: 220–44. https://doi.org/10.1016/j.jfineco.2023.04.002.

Engsted, T., and T. Q. Pedersen. 2010. “The Dividend–Price Ratio Does Predict Dividend Growth: International Evidence.” *Journal of Empirical Finance* 17: 585–605. https://doi.org/10.1016/j.jempfin.2010.01.003.

Fama, E. F., and K. R. French. 1987. "Commodity Futures Prices: Some Evidence on Forecast Power, Premiums, and the Theory of Storage." *Journal of Business* 60: 55–73. http://dx.doi.org/10.1086/296385.

Fernandez, V. 2020. "The Predictive Power of Convenience Yields." *Resources Policy* 65: 101532. https://doi.org/10.1016/j.resourpol.2019.101532.

Figuerola-Ferretti, I., C. L. Gilbert, and J. R. McCrorie. 2015. "Testing for Mild Explosivity and Bubbles in LME Non-Ferrous Metals Prices." *Journal of Time Series Analysis* 36: 763–82. https://doi.org/10.1111/jtsa.12121.

Gao, L., S. Hitzemann, I. Shaliastovich, and L. Xu. 2022. "Oil Volatility Risk." *Journal of Financial Economics* 144: 456–91. https://doi.org/10.1016/j.jfineco.2021.08.016.

Gargano, A., and A. Timmermann. 2014. "Forecasting Commodity Price Indexes using Macroeconomic and Financial Predictors." *International Journal of Forecasting* 30: 825–43. https://doi.org/10.1016/j.ijforecast.2013.09.003.

Goldstein, I., and L. Yang. 2022. "Commodity Financialization and Information Transmission." *The Journal of Finance* 77: 2613–67. https://doi.org/10.1111/jofi.13165.

Golez, B., and P. Koudijs. 2018. "Four Centuries of Return Predictability." *Journal of Financial Economics* 127: 248–63. https://doi.org/10.1016/j.jfineco.2017.12.007.

Han, M. 2023. "Commodity Momentum and Reversal: Do They Exist, and If So, Why?" *Journal of Futures Markets* 43: 1204–37. https://doi.org/10.1002/fut.22424.

Hazelkorn, T. M., T. J. Moskowitz, and K. Vasudevan. 2023. "Beyond Basis Basics: Liquidity Demand and Deviations from the Law of One Price." *The Journal of Finance* 78: 301–45. https://doi.org/10.1111/jofi.13198.

He, C., C. Jiang, and M. Molyboga. 2019. "Risk Premia in Chinese Commodity Markets." *Journal of Commodity Markets* 15: 100075. https://doi.org/10.1016/j.jcomm.2018.09.003.

Hevia, C., I. Petrella, and M. Sola. 2018. "Risk Premia and Seasonality in Commodity Futures." *Journal of Applied Econometrics* 33: 853–73. https://doi.org/10.1002/jae.2631.

Hong, H., and M. Yogo. 2012. "What Do Futures Market Interests Tell Us About the Macroeconomy and Asset Prices?" *Journal of Financial Economics* 105: 473–90.

Jacks, D. S., and M. Stuermer. 2020. "What Drives Commodity Price Booms and Busts?" *Energy Economics* 85: 104035. https://doi.org/10.1016/j.eneco.2018.05.023.

Kagraoka, Y. 2016. "Common Dynamic Factors in Driving Commodity Prices: Implications of a Generalized Dynamic Factor Model." *Economic Modelling* 52: 609–17. https://doi.org/10.1016/j.econmod.2015.10.005.

Kang, W., K. G. Rouwenhorst, and K. Tang. 2020. "A Tale of Two Premiums: The Role of Hedgers and Speculators in Commodity Futures Markets." *The Journal of Finance* 75: 377–417. https://doi.org/10.1111/jofi.12845.

Kaniel, R., G. Saar, and S. Titman. 2008. "Individual Investor Trading and Stock Returns." *The Journal of Finance* 63: 273–310. https://doi.org/10.1111/j.1540-6261.2008.01316.x.

Karali, B., and O. A. Ramirez. 2014. "Macro Determinants of Volatility and Volatility Spillover in Energy Markets." *Energy Economics* 46: 413–21. https://doi.org/10.1016/j.eneco.2014.06.004.

Kocagil, A. E. 2004. "Optionality and Daily Dynamics of Convenience Yield Behavior: An Empirical Analysis." *Journal of Financial Research* 27: 143–58. https://doi.org/10.1111/j.1475-6803.2004.00082.x.

Lutzenberger, F. T. 2014. "The Predictability of Aggregate Returns on Commodity Futures." *Review of Financial Economics* 23: 120–30. https://doi.org/10.1016/j.rfe.2014.02.001.

Maio, P., and P. Santa-Clara. 2015. "Dividend Yields, Dividend Growth, and Return Predictability in the Cross Section of Stocks." *Journal of Financial and Quantitative Analysis* 50: 33–60. https://doi.org/10.1017/S0022109015000058.

Milonas, N. T., and T. Henker. 2001. "Price Spread and Convenience Yield Behaviour in the International Oil Market." *Applied Financial Economics* 11: 23–36. https://doi.org/10.1080/09603100150210237.

Mirantes, A. G., J. Población, and G. Serna. 2013. "The Stochastic Seasonal Behavior of Energy Commodity Convenience Yields." *Energy Economics* 40: 155–66. https://doi.org/10.1016/j.eneco.2013.06.011.

Narayan, P. K., S. Narayan, and S. S. Sharma. 2013. "An Analysis Of Commodity Markets: What Gain for Investors?" *Journal of Banking & Finance* 37: 3878–89. https://doi.org/10.1016/j.jbankfin.2013.07.009.

Newey, W. K., and K. D. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55: 703–8. https://doi.org/10.2307/1913610.

Pan, W.-F. 2018. "Sentiment and Asset Price Bubble in the Precious Metals Markets." *Finance Research Letters* 26: 106–11. https://doi.org/10.1016/j.frl.2017.12.012.

Patton, A. J., and B. M., Weller 2022. "Risk Price Variation: the Missing Half of Empirical Asset Pricing." *The Review of Financial Studies* 35: 5127–84. https://doi.org/10.1093/rfs/hhac012.

Pindyck, R. S. 1993. "The Present Value Model of Rational Commodity Pricing." *The Economic Journal* 103: 511–30. https://doi.org/10.2307/2234529.

Polimenis, V., and I. M. Neokosmidis. 2016. "The Modified Dividend–Price Ratio." *International Review of Financial Analysis* 45: 31–8. https://doi.org/10.1016/j.irfa.2016.01.019.

Prokopczuk, M., A. Stancu, and L. Symeonidis. 2019. "The Economic Drivers of Commodity Market Volatility." *Journal of International Money and Finance* 98: 102063. https://doi.org/10.1016/j.jimonfin.2019.102063.

Ready, M. J., and R. C. Ready. 2022. "Order Flows and Financial Investor Impacts in Commodity Futures Markets." *The Review of Financial Studies* 35: 4712–55. https://doi.org/10.1093/rfs/hhac008.

Sharma, S., and D. Escobari. 2018. "Identifying Price Bubble Periods in the Energy Sector." *Energy Economics* 69: 418–29. https://doi.org/10.1016/j.eneco.2017.12.007.

Stambaugh, R. F. 1999. "Predictive Regressions." *Journal of Financial Economics* 54: 375–421. https://doi.org/10.1016/S0304-405X(99)00041-0.

Symeonidis, L., M. Prokopczuk, C. Brooks, and E. Lazar. 2012. "Futures Basis, Inventory and Commodity Price Volatility: An Empirical Analysis." *Economic Modelling* 29: 2651–63. https://doi.org/10.1016/j.econmod.2012.07.016.

Szymanowska, M., F. De Roon, T. Nijman, and R. Van Den Goorbergh. 2014. "An Anatomy of Commodity Futures risk Premia." *The Journal of Finance* 69: 453–82. https://doi.org/10.1111/jofi.12096.

Tang, K., and W. Xiong. 2012. "Index Investment and the Financialization of Commodities." *Financial Analysts Journal* 68: 54–74. https://doi.org/10.2469/faj.v68.n6.5.

van Huellen, S. 2019. "Price Discovery in Commodity Futures and Cash Markets with Heterogeneous Agents." *Journal of International Money and Finance* 95: 1–13. https://doi.org/10.1016/j.jimonfin.2019.03.003.

Wang, Y., L., Liu, and C. Wu. 2020. "Forecasting Commodity Prices Out-Of-Sample: Can Technical Indicators Help?" *International Journal of Forecasting* 36: 666–83. https://doi.org/10.1016/j.ijforecast.2019.08.004.

Watugala, S. W. 2019. "Economic Uncertainty, Trading Activity, and Commodity Futures Volatility." *Journal of Futures Markets* 39: 921–45. https://doi.org/10.1002/fut.22018.

Whittaker, J., C. Whitehead, and M. Somers. 2005. "The Neglog Transformation and Quantile Regression for the Analysis of a Large Credit Scoring Database." *Applied Statistics* 54: 863–78. https://doi.org/10.1111/j.1467-9876.2005.00520.x.

Yang, F. 2013. "Investment Shocks and the Commodity Basis Spread." *Journal of Financial Economics* 110: 164–84. https://doi.org/10.1016/j.jfineco.2013.04.012.

Ye, M., J. Zyren, and J. Shore. 2005. "A Monthly Crude Oil Spot Price Forecasting Model using Relative Inventories." *International Journal of Forecasting* 21: 491–501. https://doi.org/10.1016/j.ijforecast.2005.01.001.

Yeo, I.-K., and R. A. Johnson. 2000. "A New Family of Power Transformations to Improve Normality or Symmetry." *Biometrika* 87: 954–9. https://doi.org/10.1093/biomet/87.4.954.

Appendix A: Equations

A.1 Commodity index construction

Assume ski_t is the total value of the commodity index or portfolio, N is the number of commodities, $w_{i,t}$ is the weight of commodity i at time t and $\sum_{i=1}^{N} w_{i,t} = 1$, $w_{i,t} = 1/N$ for the equally weighted index and $w_{i,t} = S^s_{i,t} / \sum_{i=1}^N S^s_{i,t}$ for the price-weighted index, $S^s_{i,t}$ is the scaled spot price of commodity i at time t, $F_{i,t}$ is the futures price at time t of commodity i delivered at time $t+1$, and $n_{i,t}$ is the number of commodities i at time t in the index. So, we have $w_{i,t} = \frac{S_{i,t} n_{i,t}}{V_t}$ and $n_{i,t} = \frac{w_{i,t} V_t}{S_{i,t}}$. The gross net convenience yield of the index is then defined as:

$$\begin{aligned} Dt, t \rightarrow t+1 = D_t &= \sum_{i=1}^N D_{i,t} n_{i,t} = \sum_{i=1}^N n_{i,t} \left(S_{i,t} - \frac{F_{i,t}}{(1+rf_t)} \right) \\ &= \sum_{i=1}^N w_{i,t} V_t - \sum_{i=1}^N n_{i,t} \frac{F_{i,t}}{(1+rf_t)} = V_t - \frac{\sum_{i=1}^N n_{i,t} F_{i,t}}{1+rf_t} = S_t - \frac{F_t}{1+rf_t}, \end{aligned} \tag{A1}$$

where S_t and F_t refer to the prices of the spot index and futures index at time t. Equation (A1) implies that once S_t or F_t is determined, the other one should also be determined accordingly in order to make the gross net convenience yield D_t consistent. For example, if $S_t = V_t$, then $F_t = \sum_{i=1}^N n_{i,t} F_{i,t}$. Define the growth rate of each commodity as $\Delta S_{i,t+1} = S_{i,t+1} / S_{i,t}$, we assume that the spot price of the index increases at the rate of $\Delta S_{t+1} = \sum_{i=1}^N w_{i,t} \Delta S_{i,t+1}$. Therefore, we define the commodity spot index as:

$$S_{t+1} = \Delta S_{t+1} S_t = S_t \sum_{i=1}^N w_{i,t} \frac{S_{i,t+1}}{S_{i,t}} = \frac{S_t}{V_t} \sum_{i=1}^N n_{i,t} S_{i,t+1}. \tag{A2}$$

Assume that the index or portfolio is self-financing, then $\sum_{i=1}^N n_{i,t} S_{i,t+1} = \sum_{i=1}^N n_{i,t+1} S_{i,t+1}$. In this case, equation (A2) can be rewritten as

$$S_{t+1} = \frac{S_t}{V_t} \sum_{i=1}^N n_{i,t+1} S_{i,t+1} = \frac{S_t}{V_t} V_{t+1}. \tag{A3}$$

So, we define $S_t = V_t$, and $F_t = \sum_{i=1}^N n_{i,t} F_{i,t}$. The commodity index return is defined as $R_{t+1} = S_{t+1} / \tilde{F}_t \equiv (D_{t+1} + \tilde{F}_{t+1}) / \tilde{F}_t$, with $\tilde{F}_t = \frac{F_t}{1+rf_t}$ the current value of the futures price. As in Cochrane (2008), we calculate accumulated annual yields by first calculating the yearly return, $R_{t,t+6} = R_{t+1} \cdot R_{t+2} \cdots R_{t+6}$ (with bimonthly data, 6 periods is a year) and back out the annual accumulated net convenience yield as $D_{t,t \rightarrow t+6} = (R_{t,t+6} \cdot \tilde{F}_t - \tilde{F}_{t+6})$.

A.2 Neglog normalization parameters and GMM estimation of initial value D0

We set the normalization parameter to $\alpha = 100$, which ensures that the neglog transformation is applied to "small" numbers and, as such, retains most of the properties of the standard log. The growth rate μ is derived from futures price growth as discussed in Section 2.2, where we estimate $\mu = 0.005$ for the price-weighted index and $\mu = 0.011$ for the equally weighted index. We do not know what the "true" initial values F_0 and D_0 are, since the first observation in our sample may measure these constants with added "noise," so we estimate F_0 as the sample average of $\tilde{F}_t e^{-\mu t}$. In addition, the first observation for D_t might in fact be negative in our sample and/or sub subsamples, which can also be the case for the average of $D_t e^{-\mu t}$, and moreover, the analysis dictates that the growth rate μ applied to detrending D_t is based on F_t.

The parameter D_0 is therefore estimated with the generalized method of moments (GMMs). The moment conditions of the GMM estimation aim at matching the volatility of the yield growth rate based on the neglog with the volatility of the standard log yield growth rate, using only the observations of D_t that are positive. In addition, the moment conditions also try to match the volatility of the neglog percentage yield with the volatility of the log percentage yield for positive values of D_t.

We estimate D_0 with the GMM procedure as follows. The moments are as follows:

$$E\left(\mathrm{nl}_\alpha(\hat{D}_{t+1}) - \mathrm{nl}_\alpha(\hat{D}_t) + \mu - \lambda_1\right) = 0, \tag{A4}$$

$$E\left(\ln(D_{t+1}) - \ln(D_t) - \lambda_2\right) = 0, \tag{A5}$$

$$E\left(\left(\mathrm{nl}_\alpha(\hat{D}_{t+1}) - \mathrm{nl}_\alpha(\hat{D}_t) + \mu - \lambda_1\right)^2 - \left(\ln(D_{t+1}) - \ln(D_t) - \lambda_2\right)^2\right) = 0, \tag{A6}$$

$$E\left(\mathrm{nl}_\alpha(\hat{D}_t) + \ln(\breve{D}\ ) - \ln(\hat{F}_t) - \lambda_3\right) = 0, \tag{A7}$$

$$E\left(\ln(D_t) - \ln(\tilde{F}_t) - \lambda_4\right) = 0, \tag{A8}$$

$$E\left(\left(\mathrm{nl}_\alpha(\hat{D}_t) + \ln(\breve{D}\ ) - \ln(\hat{F}_t) - \lambda_3\right)^2 - \left(\ln(D_t) - \ln(\tilde{F}_t) - \lambda_4\right)^2\right) = 0, \tag{A9}$$

where $\hat{D}_t = (D_t/D_0)e^{-\mu t}$ and the parameter μ is fixed and equal to the average log growth rate of $\tilde{F}_t$.

Appendix B. Tables

Table B1. Description of the commodities in the dataset.

This table reports a detailed description of the dataset comprised of twenty-three commodities selected from the CRB. For each commodity, we report its name, root symbol (RS), the exchange on which the futures contracts are traded, the available delivery month, the delivery month used to construct the commodity spot and futures prices, and the initial date of the price series.

			Delivery month		
Commodities	RS	Exchange	Available	Used	Date
Corn	ZC	CBOT	3 5 7 9 12	3 5 7 9 12	1959 07
Soybean	ZS	CBOT	1 3 5 7 8 9 11	1 3 5 7 9 11	1959 07
Soybean oil	ZL	CBOT	1 3 5 7 8 9 10 12	1 3 5 7 9 12	1959 07
Soybean meal	ZM	CBOT	1 3 5 7 8 9 10 12	1 3 5 7 9 12	1960 09
Wheat	ZW	CBOT	3 5 7 9 12	3 5 7 9 12	1959 09
Oats	ZO	CBOT	3 5 7 9 12	3 5 7 9 12	1959 09
Rough rice	ZR	CBOT	1 3 5 7 9 11	1 3 5 7 9 11	1989 05
Coffee	KC	ICEUS	3 5 7 9 12	3 5 7 9 12	1972 11
Cocoa	CC	ICEUS	3 5 7 9 12	3 5 7 9 12	1959 11
Orange juice	OJ	ICEUS	1 3 5 7 9 11	1 3 5 7 9 11	1967 09
Live cattle	LE	CME	2 4 6 8 10 12	2 4 6 8 10 12	1965 01
Feeder cattle	GF	CME	1 3 4 5 8 9 10 11	1 3 5 8 9 11	1977 01
Lean hogs	HE	CME	2 4 5 6 7 8 10 12	2 4 6 8 10 12	1968 11
Lumber	LS	CME	1 3 5 7 9 11	1 3 5 7 9 11	1970 01
Cotton	CT	ICEUS	3 5 7 10 12	3 5 7 10 12	1960 09
Silver	SI	COMEX	3 5 7 9 12	3 5 7 9 12	1967 07
Copper	HG	COMEX	1 3 5 7 9 12	1 3 5 7 9 12	1959 09
Gold	GC	COMEX	2 4 6 8 10 12	2 4 6 8 10 12	1975 01
Gasoil	LF	ICE	ALL	1 3 5 7 9 11	1986 07
Natural gas	NG	NYMEX	ALL	1 3 5 7 9 11	1990 05
Crude oil	CL	NYMEX	ALL	1 3 5 7 9 11	1983 05
Gasoline	RB	NYMEX	ALL	1 3 5 7 9 11	1985 01
Heating oil	HO	NYMEX	ALL	1 3 5 7 9 11	1979 01

Table B2. Summary statistics of neglog and log variables for the US stock market.
The US stock market data are at annual frequency from December 1926 to December 2023, which are available from the Center for Research in Security Prices. The neglog dividend growth is defined as $\mathrm{nl}_\alpha(\hat{D}_{t+1}) - \mathrm{nl}_\alpha(\hat{D}_t) + \mu$, the log dividend growth is defined as $\ln(D_{t+1}/D_t)$, the neglog dividend yield is defined as $\mathrm{nl}_\alpha(\hat{D}_t) + \ln(D_0) + \mu t - \ln(P_t)$, the log dividend yield is defined as $\ln(D_t) - \ln(P_t)$, where D_t are aggregate stock dividends; the parameter values are $\alpha = 100$, $\mu = 0.047$, and $D_0 = 3.35$.

Variable	Obs	Mean	Std. Dev.	Min	Max
Neglog dividend growth	97	0.047	0.149	−0.456	0.381
Log dividend growth	97	0.047	0.141	−0.368	0.370
Neglog dividend yield	98	−2.362	0.452	−3.491	−1.560
Log dividend yield	98	−3.399	0.447	−4.495	−2.626

Table B3. Summary statistics of neglog and log variables for strictly positive net convenience yields.
This table shows the neglog and log variables for a supsample of data points that exhibit a strictly positive net convenience yield. The neglog yield growth is defined as $\mathrm{nl}_\alpha(\hat{D}_{t+1}) - \mathrm{nl}_\alpha(\hat{D}_t) + \mu$, the log yield growth is defined as $\ln(D_{t+1}/D_t)$, the neglog percentage yield is defined as $\mathrm{nl}_\alpha(\hat{D}_t) - \ln(\hat{F}_t)y_0$, and the log percentage yield is defined as $\ln(D_t) - \ln(\tilde{F}_t)$.

Variable	Obs	Mean	Std. Dev.	Min	Max
Panel A: Price-weighted index					
Neglog yield growth	204	0.099	1.030	−3.820	4.053
Log yield growth	204	0.207	1.104	−4.813	5.013
Neglog percentage yield	204	−1.380	0.837	−2.539	2.679
Log percentage yield	204	−2.701	0.803	−7.441	−1.092
Panel B: Equally weighted index					
Neglog dividend growth	184	0.132	1.310	−4.769	4.438
Log dividend growth	184	0.160	1.323	−4.675	4.116
Neglog dividend yield	184	−1.573	1.043	−3.367	2.988
Log dividend yield	184	−3.005	1.018	−8.276	−1.384

Table B4. Error covariance and correlation matrices of the VARs.
This table shows the error covariances and correlations of the VAR regression (14). Cov_ε and $Corr_\varepsilon$ are the covariance and correlation matrix of the error terms.

	$\mathbf{Cov}_\varepsilon$			$\mathbf{Corr}_\varepsilon$		
	ε^r	ε^d	ε^y	ε^r	ε^d	ε^y
Panel A: Price-weighted index						
ε^r	0.026			1		
ε^d	0.075	0.610		0.600	1	
ε^y	0.054	0.574	0.559	0.449	0.984	1
Panel B: Equally weighted index						
ε^r	0.025			1		
ε^d	0.096	1.180		0.559	1	
ε^y	0.075	1.135	1.111	0.450	0.992	1

Table B5. Regression results of the classic and generalized Campbell–Shiller decomposition for strictly positive net convenience yields.

This table shows the regression results for the classic and generalized Campbell–Shiller decomposition based on a subsample of seventeen commodities with 4-year holding-period returns that exhibit strictly positive net convenience yields. $h=1$ refers to a 4-year horizon. For commodity returns, the coefficients are the estimates of $r_{t+1}=a_r+\beta_r y_t+\varepsilon^r_{t+1}$. For the yield growth, the coefficients are the estimates of $\Delta d_{t+1}=a_d+\beta_d y_t+\varepsilon^d_{t+1}$. For the bubble, the coefficients are the estimates of $y_{t+1}=a_y+\beta_y y_t+\varepsilon^y_{t+1}$. When the horizon is ∞, the coefficients are inferred from the coefficients of 1-year horizons: $\beta^{lr}_r=\frac{\beta_r}{1-\rho\beta_y}$, $\beta^{lr}_d=\frac{\beta_d}{1-\rho\beta_y}$, $\beta^{lr}_y=\lim_{h\to\infty}\beta^h_y=0$, $\rho=0.734$ for the classic Campbell–Shiller decomposition. For the generalized Campbell–Shiller decomposition, $\mu=0.004$, $D_0=59.603$, $F_0=147.8$, and $\rho=0.713$. The t-statistics of these regressions are in parentheses and are corrected with the Newey–West correction. *, **, and *** denote statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

	Return		Yield growth		Bubble				
h	$\beta_r^{(h)}$	R_r^2	$\beta_d^{(h)}$	R_d^2	$\beta_y^{(h)}$	R_y^2	$\rho^h\beta_y^{(h)}$	$\beta_r^{(h)}-\beta_d^{(h)}+\rho^h\beta_y^{(h)}$	N
Panel A: Log variables									
1	0.313**	0.186	−0.556*	0.075	0.176	0.010	0.129	1.00	209
	(2.53)		(−1.90)		(0.73)				
∞	0.360		−0.638		0		0	1.00	
Panel B: Neglog variables									
1	0.312***	0.156	−0.592**	0.148	0.128	0.015	0.091	1.00	209
	(2.62)		(−2.37)		(0.65)				
∞	0.343		−0.652		0		0	0.99	

Appendix C. Figures

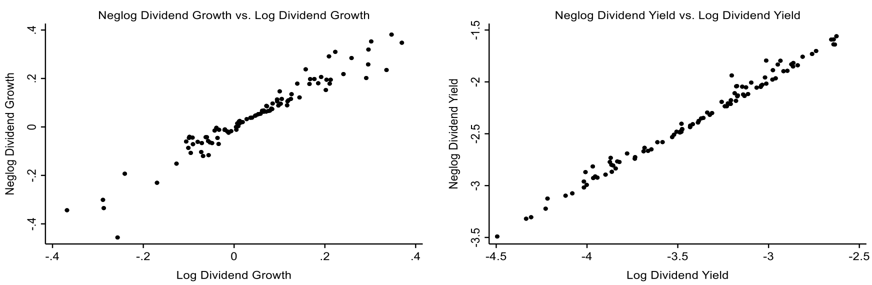

Figure C.1. Scatter plots of neglog versus log variables for US stocks.

US stock market data are at annual frequency from December 1926 to December 2023, and come from the Center for Research in Security Prices. The neglog dividend growth is defined as $\mathrm{nl}_\alpha(\hat{D}_{t+1})-\mathrm{nl}_\alpha(\hat{D}_t)+\mu$, the log dividend growth is defined as $\ln(D_{t+1}/D_t)$, the neglog dividend yield is defined as $\mathrm{nl}_\alpha(\hat{D}_t)+\ln(D_0)+\mu t-\ln(P_t)$, and the log dividend yield is defined as $\ln(D_t)-\ln(P_t)$, where D_t is the dividend of a stock, $\alpha=100$, $\mu=0.047$, and $D_0=3.35$. We do not detrend P_t here since we are only comparing variables and not applying the return decomposition—the variables are identical for any value of P_0.