Dynamic Trading with Predictable Returns and Transaction Costs

February 2012

Abstract

We derive a closed-form optimal dynamic portfolio policy when trading is costly and security returns are predictable by signals with different mean-reversion speeds. The optimal strategy is characterized by two principles: 1) aim in front of the target and 2) trade partially towards the current aim. Specifically, the optimal updated portfolio is a linear combination of the existing portfolio and an “aim portfolio,” which is a weighted average of the current Markowitz portfolio (the moving target) and the expected Markowitz portfolios on all future dates (where the target is moving). Intuitively, predictors with slower mean reversion (alpha decay) get more weight in the aim portfolio. We implement the optimal strategy for commodity futures and find superior net returns relative to more naive benchmarks.
Active investors and asset managers — such as hedge funds, mutual funds, and proprietary traders — try to predict security returns and trade to profit from their predictions. Such dynamic trading often entails significant turnover and trading costs. Hence, any active investor must constantly weigh the expected benefit of trading against its costs and risks. An investor often uses different return predictors, e.g., value and momentum predictors, and these have different prediction strengths and mean-reversion speeds, or, said differently, different “alphas” and “alpha decays.” The alpha decay is important because it determines how long time the investor can enjoy high expected returns and, therefore, affects the trade-off between returns and transactions costs. For instance, while a momentum signal may predict that the IBM stock return will be high over the next month, a value signal might predict that Cisco will perform well over the next year.

This paper addresses how the optimal trading strategy depends on securities’ current expected returns, the evolution of expected returns in the future, their risks and correlations, and their trading costs. We present a closed-form solution for the optimal dynamic portfolio strategy, giving rise to two principles: 1) aim in front of the target and 2) trade partially towards the current aim.

To see the intuition for these portfolio principles, note that the investor would like to keep his portfolio close to the optimal portfolio in the absence of transaction cost, which we call the “Markowitz portfolio.” The Markowitz portfolio is a moving target since the return-predicting factors change over time. Due to transaction costs, it is obviously not optimal to trade all the way to the target all the time. Hence, transaction costs make it optimal to reduce the speed of trading and, interestingly, to modify the aim, not trading directly towards the current Markowitz portfolio. Indeed, the optimal strategy is to trade towards an “aim portfolio” that is a weighted average of the current tangency portfolio (the moving target) and the expected tangency portfolios on all future dates (where the target is moving).

While new to finance, these portfolio principles have close analogues in other fields such as the guidance of missiles towards moving targets, shooting, and sports. For example, related
dynamic programming principles are used to guide missiles to an enemy airplane in so-called “lead homing” systems. Similarly, hunters are reminded to “lead the duck” when aiming their weapon.\footnote{We thank Kerry Back for this analogy.} The most famous example from the sports world is perhaps the following quote from the “great one”:

“A great hockey player skates to where the puck is going to be, not where it is.”

— Wayne Gretzky

Another way to state our portfolio principle is that the best new portfolio is a combination of 1) the current portfolio (to reduce turnover), 2) the Markowitz portfolio (to partly get the best current risk-return trade-off), and 3) the expected optimal portfolio in the future (a dynamic effect).

Figure 1 illustrates this natural trading rule. The solid line illustrates the the expected path a the Markowitz portfolio, starting with large positions in both security 1 and 2, and gradually converging towards its long-term mean (e.g., the market portfolio). The aim portfolio is a weighted-average of the current and future Markowitz portfolios so it lies in the “convex hull” of the solid line or, equivalently, between the current Markowitz portfolio and the expected aim portfolio next period. The optimal new position is achieved by trading partially towards this aim portfolio.

In this example, curve in the solid lines means that the Markowitz position in security 1 decays more slowly as the predictor that currently “likes” security 1 is more persistent. Therefore, the aim portfolio has a larger position in security 1, and, consequently, the optimal trade buys more shares in security 1 that it would otherwise. We show that it is in fact a more general principle that predictors with slower mean reversion (alpha decay) get more weight in the aim portfolio. An investor facing transaction costs should trade more aggressively on persistent signals than on fast mean-reverting signals: the benefits from the former accrue over longer periods, and are therefore larger.

The key role played by each return predictor’s mean reversion is an important implication of our model. It arises because transaction costs imply that the investor cannot easily change
his portfolio and, therefore, must consider his optimal portfolio both now and in the future. In contrast, absent transaction costs, the investor can re-optimize at no cost and needs to consider only the current investment opportunities (and possible hedging demands) without regard to alpha decay.

Our specification of transaction costs is sufficiently rich to allow for both purely transitory and persistent costs. Furthermore, we offer micro-foundations for each type of costs, rooted in the inventory costs of liquidity providers who take the other side of the trades. Our micro-foundation shows how the quadratic functional form arises naturally when dealers have quadratic utility and, at a deeper level, it shows how our discrete time model approaches continuous time when the trading frequency is increased.

Finally, we illustrate our results empirically in the context of commodity futures markets. We use returns over the past 5 days, 12 months, and 5 years to predict returns. The 5-day signal is quickly mean reverting (fast alpha decay), the 12-month signal mean reverts more slowly, whereas the 5-year signal is the most persistent. We calculate the optimal dynamic trading strategy taking transaction costs into account and compare its performance to the optimal portfolio ignoring transaction costs and to a class of strategies that perform static (one-period) transaction-cost optimization. Our optimal portfolio performs the best net of transaction costs among all the strategies that we consider. Its net Sharpe ratio is about 20% better than that of the best strategy among all the static strategies. Our strategy’s superior performance is achieved by trading at an optimal speed and by trading towards a aim portfolio that is optimally tilted towards the more persistent return predictors.

We also study the impulse-response of the security positions following a shock to return predictors. While the no-transaction-cost position immediately jumps up and mean reverts with the speed of the alpha decay, the optimal position increases more slowly to minimize trading costs and, depending on the alpha decay speed, may eventually become larger than the no-transaction-cost position, as the optimal position is reduced more slowly.

The paper is organized as follows. Section 2 lays out a general discrete-time model, provides a closed-form solution, and presents various related results and examples. Section 3
solves the analogous continuous-time model. Section 4 extends the model to allow for per-
sistent price impact, while Section 5 establishes the formal link between the discrete- and 
continuous-time models. Section 6 studies the model’s equilibrium implications. Section 7 
applies our framework to a trading strategy for commodity futures, and Section 8 concludes.

1 Related Literature

A large literature studies portfolio selection with return predictability in the absence of trad-
ing costs (see, e.g., Campbell and Viceira (2002) and references therein). Alpha decay plays 
no role in this literature, and nor does it play a role in the literature on optimal portfo-
lio selection with trading costs but without return predictability following Constantinides 
(1986).

This latter literature model transaction costs as proportional bid-ask spreads and relies 
on numerical solutions. Constantinides (1986) considers a single risky asset in a partial equi-
librium and studies trading-cost implications for the equity premium. Equilibrium models 
with trading costs include Amihud and Mendelson (1986), Vayanos (1998), Vayanos and 
Vila (1999), Lo, Mamaysky, and Wang (2004), Gärleanu (2009), and Acharya and Pedersen 
(2005), who also consider time-varying trading costs. Liu (2004) determines the optimal 
trading strategy for an investor with constant absolute risk aversion (CARA) and many 
independent securities with both fixed and proportional costs (without predictability). The 
assumptions of CARA and independence across securities imply that the optimal position 
for each security is independent of the positions in the other securities.

Our trade-partially-towards-the-aim strategy is qualitatively different from the optimal 
strategy with proportional or fixed transaction costs, which exhibits periods of no trading. 
Our strategy mimics a trader who is continuously “floating” limit orders close to the mid-

\footnote{Davis and Norman (1990) provide a more formal analysis of Constantinides’ model. Also, Gärleanu (2009) and Lagos and Rocheteau (2006) show how search frictions and payoff mean-reversion impact how close one trades to the static portfolio. Our continuous-time model with bounded-variation trading shares features with Longstaff (2001) and, in the context of predatory trading, by Brunnermeier and Pedersen (2005) and Carlin, Lobo, and Viswanathan (2008). See also Oehmke (2009).}
quote — a strategy that is used in practice. The trading speed (the limit orders’ “fill rate” in our analogy) depends on how large transaction costs the trader is willing to accept (i.e., on where the limit orders are placed).

In a third (and most related) strand of literature, using calibrated numerical solutions, trading costs are combined with incomplete markets by Heaton and Lucas (1996), and with predictability and time-varying investment opportunity by Balduzzi and Lynch (1999), Lynch and Balduzzi (2000), Jang, Koo, Liu, and Loewenstein (2007), and Lynch and Tan (2008). Grinold (2006) derives the optimal steady-state position with quadratic trading costs and a single predictor of returns per security. Like Heaton and Lucas (1996) and Grinold (2006), we also rely on quadratic trading costs.

A fourth strand of literature derives the optimal trade execution, treating the asset and quantity to trade as given exogenously (see, e.g., Perold (1988), Bertsimas and Lo (1998), Almgren and Chriss (2000), Obizhaeva and Wang (2006), and Engle and Ferstenberg (2007)).

Finally, quadratic programming techniques are also used in macroeconomics and other fields, and, usually, the solution comes down to algebraic matrix Riccati equations (see, e.g., Ljungqvist and Sargent (2004) and references therein). We solve our model explicitly, including the Riccati equations.

2 Discrete-Time Model

We first present the model, then solve it and provide additional results and examples.

2.1 Baseline Model

We consider an economy with $S$ securities traded at each time $t = 1, 2, 3, \ldots$. The securities’ price changes between times $t$ and $t+1$, $p_{t+1} - p_t$, are collected in an $S \times 1$ vector $r_{t+1}$ given by

$$r_{t+1} = \mu + B f_t + u_{t+1},$$

(1)
where $\mu$ is the “fair return,” e.g., from the CAPM, $u_{t+1}$ is an unpredictable zero-mean noise term with variance $\text{var}_t(u_{t+1}) = \Sigma$, $B$ is an $S \times K$ matrix of factor loadings, and $f_t$ is a $K \times 1$ vector of factors that predict returns, i.e., known to the investor already at time $t$, and evolving according to

$$
\Delta f_{t+1} = -\Phi f_t + \varepsilon_{t+1}.
$$

Here, $\Delta f_{t+1} = f_{t+1} - f_t$ is the change in the factors, $\Phi$ is a $K \times K$ matrix of mean-reversion coefficients for the factors, and $\varepsilon_{t+1}$ is the shock affecting the predictors with variance $\text{var}_t(\varepsilon_{t+1}) = \Omega$. We impose on $\Phi$ standard conditions sufficient to ensure that $f$ is stationary.\(^3\)

The interpretation of these assumptions is straightforward: the investor analyzes the securities and his analysis results in forecasts of returns. The most direct interpretation is that the investor regresses the return on security $s$ on factors $f$ that could be past returns over various horizons, valuation ratios, and other return-predicting variables, and thus estimates each variable’s ability to predict returns as given by $\beta^{sk}$ (collected in the matrix $B$). Alternatively, one can think of the factors as an analyst’s overall assessment of a security (possibly based on a range of qualitative information) and $B$ as the strength of these assessments in predicting returns.

Trading is costly in this economy and the transaction cost ($TC$) associated with trading $\Delta x_t = x_t - x_{t-1}$ shares is given by

$$
TC(\Delta x_t) = \frac{1}{2} \Delta x_t^\top \Lambda \Delta x_t,
$$

where $\Lambda$ is a symmetric positive-definite matrix measuring the level of trading costs.\(^4\) Trading costs of this form can be thought of as follows. Trading $\Delta x_t$ shares moves the (average)

---

\(^3\)We note that there are no restrictions on the positive integers $K$ and $S$ or otherwise on the factor structure. Examples 1–4 in Section 2.2 illustrate the setup by solving several concrete particular cases.

\(^4\)The assumption that $\Lambda$ is symmetric is without loss of generality. To see this, suppose that $TC(\Delta x_t) = \frac{1}{2} \Delta x_t^\top \tilde{\Lambda} \Delta x_t$, where $\tilde{\Lambda}$ is not symmetric. Then, letting $\Lambda$ be the symmetric part of $\tilde{\Lambda}$, i.e., $\Lambda = (\tilde{\Lambda} + \tilde{\Lambda}^\top)/2$, $\tilde{\Lambda}$ generates the same trading costs as $\Lambda$. 

7
price by $\frac{1}{2} \Lambda \Delta x_t$, and this results in a total trading cost of $\Delta x_t$ times the price move, which gives $TC$. Hence, $\Lambda$ (actually $\frac{1}{2} \Lambda$ for convenience) is a multi-dimensional version of Kyle’s lambda, which can also be justified by inventory considerations (e.g., Grossman and Miller (1988) or Greenwood (2005) for the multi-asset case). This transaction-cost structure arises endogenously in the model we present in Appendix A, which also shows how the parameters depend on the length of the time period. The empirical literature on price impact also lends qualitative support our to specification of transaction costs. In particular, these costs are generally found to be convex (e.g., Engle, Ferstenberg, and Russell (2008), Lillo, Farmer, and Mantegna (2003)), with some researchers actually estimating quadratic trading costs (e.g., Breen, Hodrick, and Korajczyk (2002)).

To write down the investor’s optimization problem, we introduce the term “alpha,” defined as the “abnormal” portion of the predictable return,

$$\alpha_t = B f_t.$$  \hfill (4)

The investor’s objective is to choose the dynamic trading strategy $(x_0, x_1, ...)$ to maximize the present value of all future expected alphas, penalized for risks and trading costs:

$$\max_{x_0, x_1, ...} E_0 \left[ \sum_t (1 - \rho)^t \left( x_t^\top \alpha_t - \frac{\gamma}{2} x_t^\top \Sigma x_t - \frac{1}{2} \Delta x_t^\top \Lambda \Delta x_t \right) \right],$$  \hfill (5)

where $\rho \in (0, 1)$ is a discount factor, and $\gamma$ is the risk aversion coefficient.

This objective is entirely natural. Taken literally, it captures the goal of a hedge fund manager whose mandate is to achieve “alpha” (in the usual practitioner sense). Furthermore, the model also accommodates easily an investor maximizing the total expected return, net of risk and transaction costs. To do that, one can set $\mu$ to zero in (1), and add one entry to $f$, $f_t^0 = 1$,\(^5\) and one column to $B$ given by $\beta_0^s = \mu^s$ for all $s$. In fact, this objective can be justified in a standard set-up with exponential utility for consumption and normally-distributed price changes, under certain conditions. It is easy to see that the portfolios

\(^5\)Naturally, this requires letting $\varepsilon_t^0 = 0$ for all $t$. 

8
maximizing the expected alpha, respectively the total return, are essentially the same, in
that they differ by the constant portfolio \((\gamma \Sigma)^{-1} \mu\) at all times as long as their initial values
do.

\[2.2\] Solution and Results

We solve the model using dynamic programming. We start by introducing a value function
\(V(x_{t-1}, f_t)\) measuring the value of entering period \(t\) with a portfolio of \(x_{t-1}\) securities and
observing return-predicting factors \(f_t\). The value function solves the Bellman equation:

\[
V(x_{t-1}, f_t) = \max_{x_t} \left\{ x_t^\top \alpha_t - \frac{\gamma}{2} x_t^\top \Sigma x_t - \frac{1}{2} \Delta x_t^\top \Lambda \Delta x_t + (1 - \rho)E_t[V(x_t, f_{t+1})] \right\}. \tag{6}
\]

We guess, and later verify, that the solution has a quadratic form:

\[
V(x_t, f_{t+1}) = -\frac{1}{2} x_t^\top A_{xx} x_t + x_t^\top A_{xf} f_{t+1} + \frac{1}{2} f_{t+1}^\top A_{ff} f_{t+1} + A_0, \tag{7}
\]

where we need to derive the scalar \(A_0\), the symmetric matrices \(A_{xx}\) and \(A_{ff}\), and the matrix
\(A_{xf}\).

The model in its most general form can be solved explicitly as we state in the following
proposition. The expressions for the coefficient matrices \((A_{xx}, A_{xf})\) are somewhat long, so
we leave them in the Appendix, but they become simple in the special cases discussed below,
and, in continuous time, they are relatively simple even in the most general case.

**Proposition 1** The optimal dynamic portfolio \(x_t\) is a “matrix-weighted average” of the current position and an aim portfolio:

\[
x_t = (I - \Lambda^{-1} A_{xx}) x_{t-1} + \Lambda^{-1} A_{xx} \text{aim}_t, \tag{8}
\]

with

\[
\text{aim}_t = A_{xx}^{-1} A_{xf} f_t. \tag{9}
\]
The matrix $A_{xx}$ is positive definite; $A_{xf}$ and $A_{xf}$ are stated explicitly in (B.9) and (B.15).

An alternative characterization of the optimal portfolio is as a weighted average of the current position, $x_{t-1}$, the optimal position in the absence of transaction costs, Markowitz$_t = (\gamma \Sigma)^{-1} B f_t$, and the expected target next period, $E_t(\text{aim}_{t+1}) = A^{-1}_{xx} A_{xf} (I - \Phi) f_t$:

$$x_t = \left[ \Lambda + \gamma \Sigma + (1 - \rho) A_{xx} \right]^{-1}$$

$$\times \left[ \Lambda \times x_{t-1} + \gamma \Sigma \times \text{Markowitz}_t + (1 - \rho) A_{xx} \times E_t(\text{aim}_{t+1}) \right]. \quad (10)$$

The proposition provides expressions for the optimal portfolio that are natural and relatively simple. The optimal trade $\Delta x_t$ follows directly from the proposition as

$$\Delta x_t = \Lambda^{-1} A_{xx} \left( \text{aim}_t - x_{t-1} \right). \quad (11)$$

The optimal trade is proportional to the difference between the current portfolio and the aim portfolio, and the trading speed decreases in the trading cost $\Lambda$.

We discuss the intuition behind the result further under the additional assumption that $\Lambda = \lambda \Sigma$ for some number $0 < \lambda \in \mathbb{R}$, which renders the solution even simpler. This means that the trading-cost matrix is proportional to the return variance-covariance matrix. This trading cost is natural and, in fact, implied by the model in Appendix A.1 (and that of Gârleanu, Pedersen, and Poteshman (2008)). To understand this, suppose that a dealer takes the other side of the trade $\Delta x_t$, holds this position for a period of time $h$, and “lays it off” at the end of the period. Then the dealer’s risk is $\Delta x_t^\top \Sigma h \Delta x_t$ and the trading cost is the dealer’s compensation for risk, depending on the dealer’s risk aversion reflected by $\lambda$. Under this assumption, we derive the following simple and intuitive optimal trading strategy.

**Proposition 2** When the trading cost is proportional to the amount of risk, $\Lambda = \lambda \Sigma$, the optimal new portfolio $x_t$ is a weighted average of the current position $x_{t-1}$ and a moving “target portfolio”

$$x_t = \left( 1 - \frac{a}{\lambda} \right) x_{t-1} + \frac{a}{\lambda} \text{aim}_t, \quad (12)$$
where \( \frac{a}{\lambda} < 1 \) and

\[
aim_t = (\gamma \Sigma)^{-1} B \left( I + \frac{a(1-\rho)}{\gamma} \Phi \right)^{-1} f_t
\]

(13)

\[
a = -\frac{(\gamma + \lambda \rho) + \sqrt{(\gamma + \lambda \rho)^2 + 4 \gamma \lambda (1-\rho)}}{2(1-\rho)}.
\]

(14)

The target is the optimal position in the absence of trading costs with return-predictability coefficients \( B \left( I + \frac{a(1-\rho)}{\gamma} \Phi \right)^{-1} \) instead of \( B \). An even simpler expression follows if, in addition, the matrix \( \Phi \) is diagonal: \( \Phi = \text{diag}(\phi^1, ..., \phi^K) \). In this case, the target portfolio is the optimal portfolio without trading costs and each factor \( f_t^k \) scaled depending on its alpha decay \( \phi^k \):

\[
aim_t = (\gamma \Sigma)^{-1} B \left( \frac{f_t^1}{1 + \phi^1(1-\rho)a/\gamma} \cdots \frac{f_t^K}{1 + \phi^K(1-\rho)a/\gamma} \right)^\top.
\]

(15)

Alternatively, \( x_t \) is a weighted average of the current position, \( x_{t-1} \), the optimal position in the absence of trading costs, \( \text{Markowitz}_t = (\gamma \Sigma)^{-1} B f_t \), and the expected target in the future, \( E_t(aim_{t+1}) = (\gamma \Sigma)^{-1} B \left( I + \frac{a(1-\rho)}{\gamma} \Phi \right)^{-1} (I - \Phi) f_t \):

\[
x_t = \frac{\lambda}{\lambda + \gamma(1-\rho)a} x_{t-1} + \frac{\gamma}{\lambda + \gamma(1-\rho)a} \text{Markowitz}_t + \frac{(1-\rho)a}{\lambda + \gamma(1-\rho)a} E_t(aim_{t+1}).
\]

(16)

This result provides a simple and appealing trading rule. Equation (12) states that the optimal portfolio is between the existing one and an optimal aim, where the weight on the aim \( a/\lambda \) decreases in trading costs \( \lambda \) because higher trading costs imply that one must trade more slowly. The weight on the aim increases in \( \gamma \) because a higher risk aversion means that it is more important not to let one’s position stray too far from its optimal level.

The alternative characterization (16) provides a similar intuition and comparative statics, and separates the aim into the current Markowitz optimal position without transaction costs and the expected future aim. The weight on the future aim is small if the trading cost \( \lambda \) is small (because this makes \( a \) small) or if the agent is very impatient such that \( \rho \) is close to 1. We note that while the weights on the current position \( x_{t-1} \) appear different in (12) and
(16), they are, naturally, the same. It also follows that the aim portfolio is an exponentially weighted average of the current Markowitz portfolio and the expected future Markowitz portfolios.

The optimal trading policy is illustrated in detail in Figure 2. Panel A of Figure 2 shows how the optimal first trade is derived, Panel B how the expected second trade, and Panel C the entire path of expected future trades. Let’s first understand Panel A. The solid curve is the expected path of future Markowitz portfolios. Since alphas are expected to decay to zero, the expected Markowitz portfolio converges to zero over time, meaning that the trader expects to eventually get out of these positions. In this example, asset 2 loads on a factor that decays the fastest, so the future Markowitz positions are expected to have relatively larger positions in asset 1. As a result of the general alpha decay and transaction costs, the current aim portfolio has smaller positions than the Markowitz portfolio and, as a result of the differential alpha decay, the aim portfolio loads more on in asset 1. Finally, the optimal new position is found by moving partially towards the aim.

Panel B shows that the expected next trade is towards the new aim, using the same logic as before. Panel C traces out the entire paths of expected future positions. The optimal strategy is to chase a moving aim, adjusting the aim for alpha decay and trading patiently by always edging partially towards the aim.

The optimal trading is simpler yet under the additional (and rather standard) assumption that the dynamics of each factor $f_k$ depend only on its own level (not the level of the other factors), that is, $\Phi = \text{diag} (\phi^1, \ldots, \phi^K)$ is diagonal, so that Equation (2) simplifies to scalars:

$$
\Delta f_{t+1}^k = -\phi^k f_t^k + \epsilon_{t+1}^k. 
$$

The resulting aim portfolio is very similar to the optimal portfolio without transaction costs $(\gamma \Sigma)^{-1} B f_t$. The transaction costs imply first that one optimally only trades part of the way towards the aim, and, second, that the aim down-weights each return-predicting factor more the higher is its alpha decay $\phi^k$. Down-weighting factors reduces the size of the position, and, more importantly, changes the relative importance of the different factors. This feature
is also seen in Figure 2. Note that, if the alphas of the two assets decayed equally fast, then the Markowitz portfolio would be expected to move linearly towards the origin. The convex shape of the path of expected future Markowitz portfolios, then, indicates that the factors on which asset 2 depends most heavily decay faster than those on which asset 1 does. The aim also downweights the faster decaying factors, and thus asset 2; graphically, the aim lies below the line joining the Markowitz portfolio with the origin. Naturally, giving more weight to the more persistent factors means that the investor trades towards a portfolio that not only has a high alpha now, but also is expected to have a high alpha for a longer time in the future.

We next provide a few examples.

**Example 1: Timing a single security**

An interesting and simple case is when there is only one security. This occurs when an investor is timing his long or short view of a particular security or market. In this case, the assumption that $\Lambda = \lambda \Sigma$ from Proposition 2 is without loss of generality since all parameters are scalars. In the scalar case, we use the notation $\Sigma = \sigma^2$ and $B = (\beta^1, ..., \beta^K)$. Assuming that $\Phi$ is diagonal, we can apply Proposition 2 directly to get the optimal timing trade:

$$x_t = \left(1 - \frac{a}{\lambda}\right) x_{t-1} + \frac{a}{\lambda \gamma \sigma^2} \sum_{i=1}^{K} \frac{1}{1 + \phi_i (1 - \rho) a / \gamma \beta_i^2} f_t.$$  \hfill (18)

**Example 2: Relative-value trades based on security characteristics**

It is natural to assume that the agent uses certain characteristics of each security to predict its returns. Hence, each security has its own return-predicting factors (whereas, in the general model above, all the factors could influence all the securities). For instance, one can imagine that each security is associated with a value characteristic (e.g., its own book-to-market) and a momentum characteristic (its own past return). In this case, it is natural to let the
alpha for security \( s \) be given by

\[
\alpha_t^s = \sum_i \beta_i f_t^{i,s},
\]

where \( f_t^{i,s} \) is characteristic \( i \) for security \( s \) (e.g., IBM’s book-to-market) and \( \beta_i \) be the predictive ability of characteristic \( i \) (i.e., how book-to-market translates into future expected return, for any security), which is the same for all securities \( s \). Further, we assume that characteristic \( i \) has the same mean-reversion speed for each security, that is, for all \( s \),

\[
\Delta f_{t+1}^{i,s} = -\phi_i f_t^{i,s} + \varepsilon_{t+1}^{i,s}.
\]

We collect the current values of characteristic \( i \) for all securities in a vector \( f_t^i = (f_t^{i,1}, \ldots, f_t^{i,S})^\top \), e.g., the book-to-market of security 1, book-to-market of security 2, etc.

This setup based on security characteristics is a special case of our general model. To map it into the general model, we stack all the various characteristic vectors on top of each other into \( f \):

\[
f_t = \begin{pmatrix}
f_1^t \\
\vdots \\
f_I^t
\end{pmatrix}.
\]

Further, we let \( I_{S \times S} \) be the \( S \)-by-\( S \) identity matrix and can express \( B \) using the kronecker product:

\[
B = \beta^\top \otimes I_{S \times S} = \begin{pmatrix}
\beta^1 & 0 & 0 & \beta^I & 0 & 0 \\
0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\
0 & 0 & \beta^1 & 0 & 0 & \beta^I
\end{pmatrix}.
\]

Thus, \( \alpha_t = B f_t \). Also, let \( \Phi = \text{diag}(\phi \otimes 1_{S \times 1}) = \text{diag}(\phi^1, \ldots, \phi^1, \ldots, \phi^I, \ldots, \phi^I) \). With these definitions, we apply Proposition 2 to get the optimal characteristic-based relative-value
trade as

\[
x_t = \left(1 - \frac{a}{\lambda}\right)x_{t-1} + \frac{a}{\lambda}(\gamma \Sigma)^{-1} \sum_{i=1}^{I} \frac{1}{1 + \phi^i(1 - \rho)\alpha_i / \gamma} \beta^i f_t^i.
\]

**Example 3: Static model**

When the investor completely discounts the future, i.e., \(\rho = 1\), he only cares about the current period and the problem is static. The investor simply solves

\[
\max_{x_t} x_t^\top \alpha_t - \frac{\gamma}{2} x_t^\top \Sigma x_t - \frac{\lambda}{2} \Delta x_t^\top \Sigma \Delta x_t
\]

with a solution that specializes Proposition 2:

\[
x_t = \frac{\lambda}{\gamma + \lambda} x_{t-1} + \frac{\gamma}{\gamma + \lambda} (\gamma \Sigma)^{-1} \alpha_t.
\]

To recover the optimal dynamic weight on the current position \(x_{t-1}\) from (16), one must lower the trading cost \(\lambda\) to \(\frac{1}{1+(1-\rho)\alpha/\gamma}\lambda\) to account for the future benefits of the position. Alternatively, one can increase risk aversion, or do some combination.

Interestingly, however, with multiple return-predicting factors, no choice of risk aversion \(\gamma\) and trading cost \(\lambda\) recovers the dynamic solution. This is because the static solution treats all factors the same, while the dynamic solution gives more weight to factors with slower alpha decay. We show empirically in Section 7 that even the best choice of \(\gamma\) and \(\lambda\) in a static model may perform significantly worse than our dynamic solution.

To recover the dynamic solution in a static setting, one must change not just \(\gamma\) and \(\lambda\), but additionally the alphas \(\alpha_t = B f_t\) by changing \(B\) as described in Proposition 2.

**Example 4: Today’s first signal is tomorrow’s second signal**

Suppose that the investor is timing a single market using each of the several past daily returns to predict the next return. In other words, the first signal \(f_t^1\) is the daily return for yesterday, the second signal \(f_t^2\) is the return the day before yesterday, and so on, so
that the last signal used yesterday is ignored today. In this case, the trader already knows
today what some of her signals will look like in the future. Today’s yesterday is tomorrow’s
day-before-yesterday:

\[ f_{t+1}^1 = \varepsilon_{t+1} \]
\[ f_{t+1}^k = f_t^{k-1} \quad \text{for} \ k > 1 \]

The matrix \( \Phi \) is therefore not diagonal, but has the form

\[
I - \Phi = \begin{pmatrix}
0 & 0 \\
1 & 0 \\
& \ddots \\
& & \ddots \\
0 & 1 & 0
\end{pmatrix}.
\]

Suppose for simplicity that all signals are equally important for predicting returns \( B = (\beta, \ldots, \beta) \) and use the notation \( \Sigma = \sigma^2 \). Then we can use Proposition 2 to get the optimal trading strategy

\[
x_t = \left(1 - \frac{a}{\lambda}\right) x_{t-1} + \frac{1}{\sigma^2} B((\gamma + \lambda + (1 - \rho)a)I - \lambda(1 - \rho)(I - \Phi))^{-1} f_t
\]
\[
= \left(1 - \frac{a}{\lambda}\right) x_{t-1} + \frac{\beta}{\sigma^2 (\gamma + \lambda + (1 - \rho)a)} \sum_{k=1}^{K} (1 - z^{K+1-k}) f_t^k
\]
\[
= \left(1 - \frac{a}{\lambda}\right) x_{t-1} + \frac{\beta(\lambda - a)}{\lambda^2 \sigma^2} \sum_{k=1}^{K} (1 - z^{K+1-k}) f_t^k
\]

where \( z = \frac{\lambda(1-\rho)}{\gamma + \lambda + (1-\rho)a} < 1 \). Hence, the optimal portfolio gives the largest weight to the first signal (yesterday’s return), the second largest to the second signal, and so on. This is intuitive, since the first signal will continue to be important the longest, the second signal the second longest, and so on.
3 Continuous-Time Model

We next present the continuous-time version of our model. The continuous-time model is convenient since it has an even simpler solution and, therefore, it constitutes a useful workhorse model for applications — e.g., our equilibrium analysis.

The securities have prices $p$ with dynamics

$$dp_t = (\mu + \alpha_t) \, dt + du_t, \quad \text{(26)}$$

where, as before, $\mu$ is the “fair return,” the random “noise” $u$ is a martingale (e.g., a Brownian motion) with instantaneous variance covariance matrix $\text{var}_t(du_t) = \Sigma dt$, and the predictable return $\alpha$ is given by

$$\alpha_t = Bf_t, \quad \text{(27)}$$

$$df_t = -\Phi f_t \, dt + d\varepsilon_t, \quad \text{(28)}$$

The vector $f$ contains the factors that predict returns, $B$ contains the factor loadings, $\Phi$ is the matrix of mean-reversion coefficients, and the noise term $\varepsilon$ is a martingale (e.g., a Brownian motion) with instantaneous variance-covariance matrix $\text{var}_t(d\varepsilon_t) = \Omega dt$.

The agent chooses his trading intensity $\tau_t \in \mathbb{R}^S$, which determines the rate of change$^6$ of his position $x_t$:

$$dx_t = \tau_t \, dt. \quad \text{(29)}$$

The cost per time unit of trading $\tau_t$ shares per time unit is

$$TC(\tau_t) = \frac{1}{2} \tau_t^T \Lambda \tau_t \quad \text{(30)}$$

$^6$We only consider smooth portfolio policies because discrete jumps in positions or quadratic variation would be associated with infinite trading costs in our setting. E.g., if the agent trades $n$ shares over a time period of $\Delta_t$, then the cost is $\int_0^{\Delta_t} TC\left(\frac{n}{\Delta_t}\right) dt = \frac{1}{2} \Lambda \left(\frac{n}{\Delta_t}\right)^2$, which approaches infinity as $\Delta_t$ approaches 0.
and the investor chooses his optimal trading strategy to maximize the present value of the future stream of alphas, penalized for risk and trading costs:

$$\max_{(\tau_s)_{s \geq t}} E_t \int_t^\infty e^{-\rho(s-t)} \left( x_s^\top \alpha_s - \frac{\gamma}{2} x_s^\top \Sigma x_s - \frac{1}{2} \tau_s^\top \Lambda \tau_s \right) ds. \quad (31)$$

The value function $V(x, f)$ of the investor solves the Hamilton-Jacoby-Bellman (HJB) equation

$$\rho V = \sup_\tau \left\{ x^\top B f - \gamma \frac{1}{2} x^\top \Sigma x + \frac{1}{2} \tau^\top \Lambda \tau + \frac{\partial V}{\partial x} \tau + \frac{\partial V}{\partial f} (-\Phi f) + \frac{1}{2} \text{tr} \left( \Omega \frac{\partial^2 V}{\partial f \partial f^\top} \right) \right\}. \quad (32)$$

Maximizing this expression with respect to the trading intensity results in

$$\tau = \Lambda^{-1} \frac{\partial V^\top}{\partial x}. \quad (33)$$

It is natural to conjecture a quadratic form for the value function:

$$V(x, f) = -\frac{1}{2} x^\top A_{xx} x + x^\top A_{xf} f + \frac{1}{2} f^\top A_{ff} f + A_0.$$ 

We verify the conjecture as part of the proof to the following proposition.

**Proposition 3** The optimal portfolio $x_t$ tracks a moving “aim portfolio” $A_{xx}^{-1} A_{xf} f_t$ with a tracking speed of $\Lambda^{-1} A_{xx}$. That is, the optimal trading intensity $\tau_t = \frac{dx}{dt}$ is

$$\tau_t = \Lambda^{-1} A_{xx} \left( A_{xx}^{-1} A_{xf} f_t - x_t \right), \quad (33)$$

where the positive definite matrix $A_{xx}$ and the matrix $A_{xf}$ are given by

$$A_{xx} = -\frac{\rho}{2} \Lambda + \Lambda^{\frac{1}{2}} \left( \gamma \Lambda^{-\frac{1}{2}} \Sigma \Lambda^{-\frac{1}{2}} + \frac{\rho}{4} I \right)^{\frac{1}{2}} \Lambda^{\frac{1}{2}} \quad (34)$$

$$\text{vec}(A_{xf}) = (\rho I + \Phi^\top \otimes I_K + I_S \otimes (A_{xx} \Lambda^{-1}))^{-1} \text{vec}(B). \quad (35)$$

As in discrete time, the optimal trading strategy has a particularly simple form when trading costs are proportional to the variance of the fundamentals:
Proposition 4 If trading costs are proportional to the amount of risk, $\Lambda = \lambda \Sigma$, then the optimal trading intensity $\tau_t = \frac{dx_t}{dt}$ is

$$\tau_t = \frac{a}{\lambda} (\text{aim}_t - x_t),$$

(36)

with

$$\text{aim}_t = (\gamma \Sigma)^{-1} B \left( I + \frac{a}{\gamma} \Phi \right)^{-1} f_t$$

(37)

$$a = -\rho \lambda + \sqrt{\rho^2 \lambda^2 + 4 \gamma \lambda}.$$  

(38)

In words, the optimal portfolio $x_t$ tracks aim$_t$ with speed $\frac{a}{\lambda}$. The tracking speed decreases with the trading cost $\lambda$ and increases with the risk-aversion coefficient $\gamma$.

If each factor’s alpha decay only depends on itself, $\Phi = \text{diag}(\phi_1, ..., \phi^K)$, then the aim is the optimal portfolio without transaction costs with each return-predicting factor $f_i$ down-weighted more the higher is the trading cost $\lambda$ and the higher is its alpha decay speed $\phi^k$:

$$\text{aim}_t = (\gamma \Sigma)^{-1} B \left( \frac{f_1}{1 + a \phi^1 / \gamma}, ..., \frac{f^K}{1 + a \phi^K / \gamma} \right)^\top.$$  

(39)

When the agent is very patient, that is, $\rho = 0$, the expressions are even simpler. The coefficient $a$ is simply $a = \sqrt{\gamma \lambda}$, and the tracking speed is $\frac{a}{\lambda} = \sqrt{\frac{\gamma}{\lambda}}$, which clearly decreases with trading costs $\lambda$ and increases with risk aversion $\gamma$.

4 Persistent Price Impact

In some cases trading may have a significant persistent price impact in addition to the transitory trading cost that we have studied so far. To study this situation, we consider an investor that can transact at a price $\tilde{p}_t = p_t + D_t$ by paying a transitory trading cost $TC$. Here, $p$ is the price without the effect of the investor’s own trading (as before), $TC$ is as before, and the new term $D_t$ captures the accumulated price distortion due to the investor’s

---

7These statements are meant to be interpreted as limits as $\rho \to 0$, to maintain the problem well defined.
trades. Trading an amount $\Delta x$ pushes prices by $C\Delta x$, and the price distortion mean reverts at a speed (or “resiliency”) $R$:

$$D_{t+1} = (I - R)D_t + C\Delta x_{t+1}. \quad (40)$$

The investor’s objective is as before (i.e., (5)), but now the securities’ alpha — here, for mathematical convenience, alpha is price appreciation net of required return $\mu$ and the exogenous unpredictable part $u$, i.e., $\alpha = \bar{p}_{t+1} - \bar{p}_t - \mu - u_{t+1}$ — incorporates both the effect of predictability of $p_{t+1}$ by the factors $f_t$ and of the predictability due to price distortions (changes in $D_t$):

$$\alpha_t = Bf_t - RD_t + C\Delta x_{t+1}. \quad (41)$$

The continuous-time formulation, on which we focus from here on for simplicity, mirrors closely (40) and (41):

$$dD_t = -RD_t \, dt + C\tau_t \, dt \quad (42)$$

$$\alpha_t = Bf_t - RD_t + C\tau_t. \quad (43)$$

The value function is now quadratic in the extended state variable $(x_t, y_t) \equiv (x_t, f_t, D_t)$ (respectively $(x_{t-1}, y_t) \equiv (x_{t-1}, f_t, D_{t-1})$, in discrete time):

$$V(x, y) = -\frac{1}{2}x^\top A_{xx}x + x^\top A_{xy}y + \frac{1}{2}y^\top A_{yy}y + A_0.$$  

We solve the HJB, respectively Bellman, equation as before.

**Proposition 5** The optimal portfolio $x_t$ tracks a moving “aim portfolio”. The optimal trad-
The optimal trading policy has a similar structure to before, but the persistent price impact changes both the speed of trading and the target portfolio, as can be seen in Equations (B.39)–(B.40), respectively (B.59)–(B.60).

5 Connection between Discrete and Continuous Time

In Appendix A, we detail a micro foundation that shows how quadratic transaction costs arise endogenously as a result of dealers’ risk aversion. Further, this model extension allows us to analyze the effect of increasing the trading frequency. Such an analysis is valuable both for providing insights into the robustness of the economic arguments and for showing how to scale parameters with the period length.

Building on this analysis, we derive the following proposition, which states that both transaction costs and the optimal strategy in discrete time tend to the ones obtaining in continuous time as the duration of a period goes to zero. Proposition 6 not only establishes that the discrete-time and continuous-time formulations are compatible with each other, but also suggests that one may use the continuous-time model as an approximation to the discrete-time model. We find in the context of the concrete setting of Section 7, where we consider daily trading in fifteen commodity futures based on three signals for each commodity, that the daily correlation between trading profits in the discrete-time daily model and the continuous-time model is 99.99%.

To facilitate the linking of the discrete- and continuous-time specifications in the following section, we make explicit here the length $\Delta_t$ of a time period.
Proposition 6 Consider the discrete-time model with both transitory and persistent transaction costs with parameters defined to depend on the time interval $\Delta t$ according to the model extension in Appendix A and summarized in Equations (A.19)–(A.26). As $\Delta t$ approaches zero, the discrete-time solution approaches the continuous-time solution: The coefficient matrices of Proposition 5 converge, $M^{\text{speed}} \rightarrow \bar{M}^{\text{speed}}$ and $M^{\text{aim}} \rightarrow \bar{M}^{\text{aim}}$, as does the trading intensity $\Delta x_t/\Delta t \rightarrow \tau_t$ for fixed state variable.

Most of our parameters are scaled in a straightforward way. Thus, as is standard, expected returns and variances are linear in the period length $\Delta t$ (to keep them at a constant annualized rate), while risk aversion is independent of $\Delta t$. Mean reversion rates and discount rates are set to keep a constant annual rate using compounding. The only non-standard parameter to scale is the quadratic transaction-cost parameter $\Lambda$, which increases linearly with the trading frequency, and this is where our micro foundation is especially helpful.

We sketch the intuition behind the proposition, focusing on transitory costs. Transaction costs arise as compensation for dealers that take the other side of the trade and hold the newly acquired position for a certain period of time $h$. While each dealer only trades $h$ time units apart, there are always dealers present — albeit fewer as $\Delta t$ decreases — since dealers arrive at different times. Each dealer’s utility cost equals his risk aversion times the dollar variance of his position over the holding period. Consequently, the endogenous $\Lambda$ equals the aggregate risk aversion of all $\Delta t/h$ dealers participating in a trade multiplied by the price variance over the dealer holding period $h$.

Consider now doubling the trading frequency (from $\Delta t^{-1}$ to $(\Delta t/2)^{-1}$). The number of dealers involved in each trade halves, and thus their (aggregate) risk aversion doubles. On the other hand, the price variance over the holding period remains the same, so that $\Lambda$ doubles. Consequently, the effect on the total transaction cost is given by a multiplicative factor of $2 \times (1/2)^2 \times 2 = 1$: the first term is due to $\Lambda$, the second to the halved position size, and the third to the fact that there are two rounds of trading during $\Delta t$.

In other words, transaction costs are stable as $\Delta t \rightarrow 0$ — i.e., to the first order, are independent of $\Delta t$. This phenomenon is entirely natural. The simple intuition is that
each dealer's inventory risk is the same irrespective of $\Delta_t$ — both the inventory size and holding period are constant. Hence, the economy behaves the same way for all $\Delta_t$ ($\Delta_t$ merely determines how frequently one observes the economy), capturing the fact that a given economy can be modeled equivalently using discrete or continuous time.

We note here that one could make different assumptions concerning the dealers’ holding periods, with different implications. In particular, if $h$ decreases with the trading-period length $\Delta_t$ — say, $h = \Delta_t$ — then transitory transaction costs vanish in the limit as $\Delta_t \to 0$.

Finally, to model persistent price impact, we consider dealers who are able to unload their inventories at a constant rate, which gives the decay rate of the price impact. The details are in Appendix A.2.

6 Equilibrium Implications

In this section we study the restrictions placed on a security’s return properties by the market equilibrium. More specifically, we consider a situation in which an investor facing transaction costs absorbs a residual supply specified exogenously and analyze the relationship implied between the characteristics of the supply dynamics and the return alpha.

For simplicity, we consider a model with one security in which $L \geq 1$ groups of (exogenously given) noise traders hold positions $z^l_t$ (net of the aggregate supply) given by

\begin{align}
\frac{dz^l_t}{dt} &= \kappa (f^l_t - z^l_t)dt \\
\frac{df^l_t}{dt} &= -\psi_t f^l_t dt + dW^l_t. \tag{46}
\end{align}

In addition, the Brownian motions $W^l_t$ satisfy $\text{var}(dW^l_t)/dt = \Omega_{ll}$. It follows that the aggregate noise-trader holding, $z_t = \sum_l z^l_t$, satisfies

\begin{align}
\frac{dz_t}{dt} &= \kappa \left( \sum_{l=1}^L f^l_t - z_t \right)dt. \tag{48}
\end{align}

We conjecture that the investor’s inference problem is as studied in Section 3, where $f$ given
by \( f \equiv (f^1, \ldots, f^L, z) \) is a linear return predictor and \( B \) is to be determined. We verify the conjecture and find \( B \) as part of Proposition 7 below.

Given the definition of \( f \), the mean-reversion matrix \( \Phi \) is given by

\[
\Phi = \begin{pmatrix}
\psi_1 & 0 & \cdots & 0 \\
0 & \psi_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
-\kappa & -\kappa & \cdots & \kappa
\end{pmatrix}.
\]

(49)

Suppose that the only other investors in the economy are the investors considered in Section 3, facing transaction costs given by \( \Lambda = \lambda \sigma^2 \). In this simple context, an equilibrium is defined as a price process and market-clearing asset holdings that are optimal for all agents given the price process. Since the noise traders' positions are optimal by assumption as specified by (46)–(47), the restriction imposed by equilibrium is that the dynamics of the price are such that, for all \( t \),

\[
x_t = -z_t
\]

(50)

\[
dx_t = -dz_t.
\]

(51)

Using (36), these equilibrium conditions lead to

\[
\frac{a}{\lambda} \sigma^{-2} B (a \Phi + \gamma I)^{-1} + \frac{a}{\lambda} e_{L+1} = -\kappa (1 - 2e_{L+1}),
\]

(52)

where \( e_{L+1} = (0, \cdots, 0, 1) \in \mathbb{R}^{L+1} \) and \( 1 = (1, \cdots, 1) \in \mathbb{R}^{L+1} \). It consequently follows that, if the investor is to hold \(-f^L_t\) at time \( t \) for all \( t \), then the factor loadings must be given by

\[
B = \sigma^2 \left[ -\frac{\lambda}{a} \kappa (1 - 2e_{L+1}) - e_{L+1} \right] (a \Phi + \gamma I)
\]

\[
= \sigma^2 \left[ -\lambda \kappa (1 - 2e_{L+1}) - ae_{L+1} \right] \left( \Phi + \frac{\gamma}{a} I \right).
\]

(53)
For \( l \leq L \), we calculate \( B_l \) further as

\[
B_l = -\sigma^2 \kappa (\lambda \psi_l + \lambda \gamma a^{-1} + \lambda \kappa - a) \\
= -\lambda \sigma^2 \kappa (\psi_l + \rho + \kappa),
\]

while

\[
B_{L+1} = \sigma^2 (\rho \lambda \kappa + \lambda \kappa^2 - \gamma).
\]

Using this, it is straightforward to see the following key equilibrium implications:

**Proposition 7** The market is in equilibrium if and only if \( x_0 = -z_0 \) and the security’s alpha is given by

\[
\alpha_t = \sum_{l=1}^{L} \lambda \sigma^2 \kappa (\psi_l + \rho + \kappa) (-f_t^l) + \sigma^2 (\rho \lambda \kappa + \lambda \kappa^2 - \gamma) z_t.
\]

The coefficients \( \lambda \sigma^2 \kappa (\psi_k + \rho + \kappa) \) are positive and increase in the mean-reversion parameters \( \psi_k \) and \( \kappa \) and in the trading costs \( \lambda \sigma^2 \). In other words, noise trader selling \( (f_t^k < 0) \) increases the alpha, and especially so if its mean reversion is faster and if the trading cost is larger.

Naturally, noise-trader selling increases the expected excess return (alpha), while noise-trader buying lowers the alpha, since the arbitrageurs need to be compensated to take the other side of the trade. Interestingly, the effect is larger when trading costs are larger and for noise-trader shocks with faster mean reversion because such shocks are associated with larger trading costs for the arbitrageurs.

### 7 Application: Dynamic Trading of Commodity Futures

In this section we illustrate our approach using data on commodity futures. We show how dynamic optimizing can improve performance in an intuitive way, and how it changes the
way new information is used.

7.1 Data

We consider 15 different liquid commodity futures, which do not have tight restrictions on the size of daily price moves (limit up/down). In particular, as seen in Table 1, we collect data on Aluminum, Copper, Nickel, Zinc, Lead, and Tin from the London Metal Exchange (LME), on Gas Oil from the Intercontinental Exchange (ICE), on WTI Crude, RBOB Unleaded Gasoline, and Natural Gas from the New York Mercantile Exchange (NYMEX), on Gold and Silver from the New York Commodities Exchange (COMEX), and on Coffee, Cocoa, and Sugar from the New York Board of Trade (NYBOT). (This excludes futures on various agriculture and livestock that have tight price limits.) We consider the sample period 01/01/1996 – 01/23/2009, for which we have data on all the commodities.9

For each commodity and each day, we collect the futures price measured in U.S. dollars per contract. For instance, if the gold price is $1,000 per ounce, the price per contract is $100,000, since each contract is for 100 ounces. Table 1 provides summary statistics on each contract’s average price, the standard deviation of price changes, the contract multiplier (e.g., 100 ounces per contract in the case of gold), and daily trading volume.

We use the most liquid futures contract of all maturities available. By always using data on the most liquid futures, we are implicitly assuming that the trader’s position is always held in these contracts. Hence, we are assuming that when the most liquid futures nears maturity and the next contract becomes more liquid, the trader “rolls” into the next contract, i.e., replaces the position in the near contract with the same position in the far contract. Given that rolling does not change a trader’s net exposure, it is reasonable to abstract from the transaction costs associated with rolling. (Traders in the real world do in fact behave like this. There is a separate roll market, which entails far smaller costs than independently selling the “old” contract and buying the “new” one.) When we compute price changes, we

---

9Our return predictors use moving averages of price data lagged up to five years, which are available for most commodities except some of the LME base metals. In the early sample when some futures do not have a complete lagged price series, we use the average of the available data.
always compute the change in price of a given contract (not the difference between the new contract and the old one), since this corresponds to an implementable return. Finally, we collect data on the average daily trading volume per contract as seen in the last column of Table 1. Specifically, we receive an estimate of the average daily volume of the most liquid contract traded electronically and outright (i.e., not including calendar-spread trades) in December 2010 from an asset manager based on underlying data from Reuters.

### 7.2 Predicting Returns and Other Parameter Estimates

We use the characteristic-based model described in Example 2 in Section 2, where each commodity characteristic is its own past returns at various horizons. Hence, to predict returns, we run a pooled panel regression:

\[
    r_{t+1}^s = 0.001 + 10.32 f_{t}^{5D,s} + 122.34 f_{t}^{1Y,s} - 205.59 f_{t}^{5Y,s} + u_{t+1}^s, \tag{57}
\]

where the left-hand side is the daily commodity price changes and the right-hand side contains the return predictors: \( f_{t}^{5D} \) is the average past five days’ price changes, divided by the past five days’ standard deviation of daily price changes, \( f_{t}^{1Y} \) is the past year’s average daily price change divided by the past year’s standard deviation, and \( f_{t}^{5Y} \) is the analogous quantity for a five-year window. Hence, the predictors are rolling Sharpe ratios over three different horizons, and, to avoid dividing by a number close to zero, the standard deviations are winsorized below the average tenth percentile of standard deviations. We estimate the regression using feasible generalized least squares and report the t-statistics in brackets.

We see that price changes show continuation at short and medium frequencies and reversal over long horizons.\(^{10}\) The goal is to see how an investor could optimally trade on this information, taking transaction costs into account. Of course, these (in-sample) re-

\(^{10}\)Erb and Harvey (2006) document 12-month momentum in commodity futures prices. Asness, Moskowitz, and Pedersen (2008) confirm this finding and also document 5-year reversals. These results are robust and hold both for price changes and returns. The 5-day momentum is less robust. For instance, for certain specifications using percent returns, the 5-day coefficient switches sign to reversal. This robustness is not important for our study due to our focus on optimal trading rather than out-of-sample return predictability.
Regression results are only available now and a more realistic analysis would consider rolling out-of-sample regressions. However, using the in-sample regression allows us to focus on the economic insights underlying our novel portfolio optimization. Indeed, the in-sample analysis allows us to focus on the benefits of giving more weight to signals with slower alpha decay, without the added noise in the predictive power of the signals arising when using out-of-sample return forecasts.

The return predictors are chosen so that they have very different mean reversion:

\[
\begin{align*}
\Delta f_{t+1}^{5D,s} &= -0.2519 f_t^{5D,s} + \varepsilon_{t+1}^{5D,s} \\
\Delta f_{t+1}^{1Y,s} &= -0.0034 f_t^{1Y,s} + \varepsilon_{t+1}^{1Y,s} \\
\Delta f_{t+1}^{5Y,s} &= -0.0010 f_t^{5Y,s} + \varepsilon_{t+1}^{5Y,s}.
\end{align*}
\]  

(58)

These mean reversion rates correspond to a 2.4-day half life for the 5-day signal, a 206-day half life for the 1-year signal, and a 700-day half life for the 5-year signal.\(^{11}\)

We estimate the variance-covariance matrix \(\Sigma\) using daily price changes over the full sample, shrinking the correlations 50% towards zero. We set the absolute risk aversion to \(\gamma = 10^{-9}\), which we can think of as corresponding to a relative risk aversion of 1 for an agent with 1 billion dollars under management. We set the time discount rate to \(\rho = 1 - \exp(-0.02/260)\) corresponding to a 2 percent annualized rate.

Finally, to choose the transaction-cost matrix \(\Lambda\), we make use of price-impact estimates from the literature. In particular, we use the estimate from Engle, Ferstenberg, and Russell (2008) that trades amounting to 1.59% of the daily volume in a stock have a price impact of about 0.10%. (Breen, Hodrick, and Korajczyk (2002) provides a similar estimate.) Further, Greenwood (2005) finds evidence that market impact in one security spills over to other securities using the specification \(\Lambda = \lambda \Sigma\), where we recall that \(\Sigma\) is the variance-covariance matrix. We calibrate \(\Sigma\) as the empirical variance-covariance matrix of price changes, where the covariance are shrunk 50% towards zero for robustness.

\(^{11}\)The half life is the time it is expected to take for half the signal to disappear. It is computed as \(\log(0.5)/\log(1 - 0.2519)\) for the 5-day signal.
We choose the scalar $\lambda$ based on the Engle, Ferstenberg, and Russell (2008) estimate by calibrating it for each commodity and then computing the mean and median across commodities. Specifically, we collect data on the trading volume of each commodity contract as seen in last column of Table 1 and then calibrate $\lambda$ for each commodity as follows. Consider, for instance, unleaded gasoline. Since gasoline has a turnover of 11,320 contracts per day and a daily price-change volatility of $1,340, the transaction cost per contract when one trades 1.59% of daily volume is $1.59\% \times 11,320 \times \lambda^{\text{Gasoline}} / 2 \times 1,340^2$, which is 0.10% of the average price per contract of $48,000 if $\lambda^{\text{Gasoline}} = 3 \times 10^{-7}$.

We calibrate the trading costs for the other commodities similarly, and obtain a median value of $5.0 \times 10^{-7}$ and a mean of $8.4 \times 10^{-7}$. There are significant differences across commodities (e.g., the standard deviation is $1.0 \times 10^{-6}$), reflecting that these estimates are based on turnover while the specification $\Lambda = \lambda \Sigma$ assumes that transaction costs depend on variances. While our model is general enough to handle transaction costs that depend on turnover (e.g., by using these calibrated $\lambda$’s in the diagonal of the $\Lambda$ matrix), we also need to estimate the spill-over effects (i.e., the off-diagonal elements). Since Greenwood (2005) provides the only estimate of these transaction-cost spill-overs in the literature using the assumption $\Lambda = \lambda \Sigma$ and since real-world transaction costs likely depend on variance as well as turnover, we stick to this specification and calibrate $\lambda$ as the median across the estimates for each commodity. Naturally, other specifications of the transaction-cost matrix would give slightly different results, but our main purpose is simply to illustrate the economic insights that we have proved in general theoretically.

We also consider a more conservative transaction cost estimate of $\lambda = 10 \times 10^{-7}$. Alternatively, this more conservative analysis can be interpreted as the trading strategy of a larger investor (i.e., we could have equivalently reduced the absolute risk aversion $\gamma$).

### 7.3 Dynamic Portfolio Selection with Trading Costs

We consider three different trading strategies: the optimal trading strategy given by Equation (23) (“optimal”), the optimal trading strategy in the absence of transaction costs
(“Markowitz”), and a number of trading strategies based on a static (i.e., one-period) transaction-cost optimization as in Equation (25) (“static optimization”). The static portfolio optimization results in trading partially towards the Markowitz portfolio (as opposed to an aim that depends on signals’ alpha decays) and we consider ten different trading speeds in seen in Table 2. Hence, under the static optimization, the updated portfolio is a weighted average of the Markowitz portfolio (with weight denoted “weight on Markowitz”) and the current portfolio.

Table 2 reports the performance of each strategy as measured by, respectively, its Gross Sharpe Ratio and its Net Sharpe Ratio (i.e., its Sharpe ratio after accounting for transaction costs). Panel A reports these numbers using our base-case transaction-cost estimate (discussed above), while Panel B uses our high transaction-cost estimate. We see that, naturally, the highest SR before transaction costs is achieved by the Markowitz strategy. The optimal and static portfolios have similar drops in gross SR due to their slower trading. After transaction costs, however, the optimal portfolio is the best, significantly better than the best possible static strategy, and the Markowitz strategy incurs enormous trading costs.

It is interesting to consider the driver of the superior performance of the optimal dynamic trading strategy relative to the best possible static strategy. The key to the out-performance is that the dynamic strategy gives less weight to the 5-day signal because of its fast alpha decay. The static strategy simply tries to control the overall trading speed, but this is not sufficient: it either incurs large trading costs due to its “fleeting” aim (because of the significant reliance on the 5-day signal), or it trades so slowly it is difficult to capture the alpha. The dynamic strategy overcomes this problem by trading somewhat fast, but trading mainly according to the more persistent signals.

To illustrate the difference in the positions of the different strategies, Figure 3 shows the positions over time of two of the commodity futures, namely Crude and Gold. We see that the optimal portfolio is a much smoother version of the Markowitz strategy, thus reducing trading costs while at the same time capturing most of the excess return. Indeed, the optimal position tends to be long when the Markowitz portfolio is long and short when
the Markowitz portfolio is short, and to be larger when the expected return is large, but moderates the speed and magnitude of trades.

7.4 Response to New Information

It is instructive to trace the response to a shock to the return predictors, namely to $\varepsilon_i^{t,s}$ in Equation (58). Figure 4 shows the responses to shocks to each return-predicting factor, namely the 5-day factor, the 1-year factor, and the 5-year factor.

The first panel shows that the Markowitz strategy immediately jumps up after a shock to the 5-day factor and slowly mean reverts as the alpha decays. The optimal strategy trades much more slowly and never accumulates nearly as large a position. Interestingly, since the optimal position also trades more slowly out of the position as the alpha decays, the lines cross as the optimal strategy eventually has a larger position than the Markowitz strategy.

The second panel shows the response to the 1-year factor. The Markowitz jumps up and decays, whereas the optimal position increases more smoothly and catches up as the Markowitz starts to decay. The third panel shows the same for the 5Y signal, except that the effects are slower and with opposite sign, since 5-year returns predict future reversals.

8 Conclusion

This paper provides a highly tractable framework for studying optimal trading strategies in light of various return predictors, risk and correlation considerations, as well as transaction costs. We derive an explicit closed-form solution for the optimal trading policy and highlight several useful and intuitive results. The optimal portfolio tracks an “aim portfolio,” which is analogous to the optimal portfolio in the absence of trading costs in its tradeoff between risk and return, but different since more persistent return predictors are weighted more heavily relative to return predictors with faster alpha decay. The optimal strategy is not to trade all the way to the aim portfolio, since this entails too high transaction costs. Instead, it is optimal to take a smoother and more conservative portfolio that moves in the direction of
the aim portfolio while limiting turnover.

Our framework constitutes a powerful tool to optimally combine various return predictors taking into account their evolution over time, decay rate, and correlation, and trading off their benefits against risks and transaction costs. Such trade-offs are at the heart of the decisions of “arbitrageurs” that help make markets efficient as per the efficient market hypothesis. Arbitrageurs’ ability to do so is limited, however, by transaction costs, and our model provides a tractable and flexible framework for the study of the dynamic implications of this limitation.

We illustrate this feature by embedding our setting in an equilibrium model with several “noise traders” who trade in and out of their positions with varying mean-reversion speeds. In equilibrium, a rational arbitrageur — with trading costs and using the methodology that we derive — needs to take the other side of these noise-trader positions to clear the market. We solve the equilibrium explicitly and show how noise trading leads to return predictability and return reversals. Further, we show that noise-trader demand that mean-reverts more quickly leads to larger return predictability. This is because a fast mean reversion is associated with high transaction costs for the arbitrageurs and, consequently, they must be compensated in the form of larger return predictability.

We implement our optimal trading strategy for commodity futures. Naturally, the optimal trading strategy in the absence of transaction costs has a larger Sharpe ratio gross of fees than our trading policy. However, net of trading costs our strategy performs significantly better, since it incurs far lower trading costs while still capturing much of the return predictability and diversification benefits. Further, the optimal dynamic strategy is significantly better than the best static strategy — taking dynamics into account significantly improves performance.

In conclusion, we provide a tractable solution to the dynamic trading strategy in a relevant and general setting that we believe to have many interesting applications.
A  Micro-Foundation for Transaction Costs

In this section we provide formal arguments for our modeling choices by presenting micro-foundations for the types of costs we consider and for their dependence on the period length. As a consequence, we also show that the discrete-time model, when specified in accordance to the micro-foundations, tends to the continuous-time model as the period length goes to 0.

A.1 Transitory Price Impact Costs

To obtain a temporary price impact of trades endogenously, we consider an economy populated by three types of investors: (i) the trader whose optimization problem we study in the paper, referred throughout this section as “the trader,” (ii) “market makers,” who act as intermediaries, and (iii) “end users,” on whom market makers eventually unload their positions as described below.

The temporary price impact is due to the market makers’ inventories. We assume that there are a mass-one continuum of market makers indexed by the set \([0, h]\) and they arrive for the first time at the market at a time equal to their index. The market operates only at discrete times \(\Delta_t\) apart,\(^\text{12}\) and the market makers trade at the first trading opportunity. Once they trade — say, at time \(t\) — market makers must spend \(h\) units of time gaining access to end users. At time \(t + h\), therefore, they unload their inventories at a price \(p_{t+h}\) described below, and rejoin the market immediately thereafter. It follows that at each trading date in the market there is always a mass \(\frac{\Delta_t}{h}\) of competing market makers that clear the market.

The price \(p\), the competitive price of end users, follows an exogenous process and corresponds to the fundamental price in the body of the paper. Market makers take this price as given and trade a quantity \(q\) to maximize a quadratic utility:

\[
\max_q \left\{ \hat{E}_t \left[ q(p_{t+h} - e^{rh} \hat{p}_t) \right] - \frac{\gamma M}{2} \text{Var}_t \left[ q(p_{t+h} - e^{rh} \hat{p}_t) \right] \right\}, \tag{A.1}
\]

\(^{12}\)We make the simplifying assumption that \(\frac{h}{\Delta_t}\) is an integer.
where \( \hat{p}_t \) is the market price at time \( t \) and \( r \) is the (continuously-compounded) risk-free rate over the horizon. \( \hat{E} \) denotes expectations under the probability measure obtained from the market makers’ beliefs using their (normalized) marginal utilities corresponding to \( q = 0 \) as Radon-Nikodym derivative. Consequently,

\[
\hat{E}_t [p_{t+h}] = e^{rh}p_t,
\]

so that the maximization problem becomes

\[
\max_q \left\{ q(p_t - \hat{p}_t) - e^{-rh} \frac{\gamma^M}{2} Var_t [qp_{t+h}] \right\}. \tag{A.2}
\]

The price \( \hat{p} \) is set so as to satisfy the market-clearing condition

\[
0 = \Delta x_t + q \frac{\Delta t}{h}. \tag{A.3}
\]

Since \( p \) is exogenous and Gaussian with variance \( V_h \) \( h \) periods ahead that can be calculated easily,\(^\text{13}\) the maximization problem yields

\[
\hat{p}_t = p_t + e^{-rh} \gamma^M V_h \frac{\Delta x_t}{\Delta t} h. \tag{A.4}
\]

Consequently, if the trader trades an amount \( \Delta x_t \), he trades at the unit price of \( p_t \) and pays an additional transaction cost of

\[
e^{-rh} \gamma^M \Delta x_t^\top V_h \frac{\Delta x_t}{\Delta t} h,
\]

which has the quadratic form posited in the body of the paper.

Two cases suggest themselves naturally when considering the choice for the holding period \( h \) as a function of \( \Delta t \). In the first case, a decreasing \( \Delta t \) is thought of as an improvement in the trading technology, attention, etc., of all market participants, and therefore \( h \) decreases as \( \Delta t \) does — in its simplest form, \( h = \Delta t \), which yields a transaction cost of the order \( \Delta t^2 \).

\(^\text{13}\)The resulting value is \( V_h = \Sigma h + B N_h \Omega N_h^\top B^\top \), where \( N_h = \int_0^h \int_u^h e^{-\Phi(t-u)} dt du = \Phi^{-1} h - \Phi^{-2} (I - e^{-\Phi h}) \) if \( \Phi \) is invertible. (Note that the first term, \( \Sigma h \), is of order \( h \), while the second of order \( h^2 \).)
Generally, as long as $h \to 0$ as $\Delta t \to 0$, the transaction costs also vanishes. The second case is that of a constant $h$: the dealers need a fixed amount of time to lay off a position regardless of the frequency with which our original traders access the market. It follows, in this case, that the price impact does not vanish as $\Delta t$ becomes small: in the continuous-time limit ($\Delta t \to 0$), the per-unit-of-time transaction cost is proportional to

$$\lim_{\Delta t \to 0} \frac{\Delta x_t^\top}{\Delta t} V_h h \frac{\Delta x_t}{\Delta t} = \tau^\top V_h h \tau,$$

as assumed in Section 3. One can therefore interpret $\Delta t$ in this case as the frequency with which the researcher observes the world, which does not impact (to the first order) equilibrium quantities — in particular, flow trades and costs.

### A.2 Persistent Price Impact

A similar model, but with a different specification of the market makers, can be used to justify a persistent price impact. Consider therefore the same model as in the previous section, but suppose now that market makers do not hold their inventories for a deterministic number $h$ of time units, but rather manage to deplete them, through trade with end users at price $p$, at a constant rate $\psi$. Thus, between two trading dates with the trader, a market maker’s inventory evolves according to

$$\Delta I_{t+1} = -\psi I_t \Delta t + q_t,$$  

where, in equilibrium,

$$q_t = \Delta x_t.$$

The market makers continue to maximize a quadratic objective:

$$\max_q \left\{ \hat{E}_t \sum_{s=t+n\Delta_t} e^{-r(s-t)} \left( \psi I_s^\top p_s \Delta_t q_s^\top \hat{p}_s - \frac{\gamma^M}{2} I_s^\top V_{\Delta_t} I_s \right) \right\},$$

as assumed in Section 3. One can therefore interpret $\Delta t$ in this case as the frequency with which the researcher observes the world, which does not impact (to the first order) equilibrium quantities — in particular, flow trades and costs.

### A.2 Persistent Price Impact

A similar model, but with a different specification of the market makers, can be used to justify a persistent price impact. Consider therefore the same model as in the previous section, but suppose now that market makers do not hold their inventories for a deterministic number $h$ of time units, but rather manage to deplete them, through trade with end users at price $p$, at a constant rate $\psi$. Thus, between two trading dates with the trader, a market maker’s inventory evolves according to

$$\Delta I_{t+1} = -\psi I_t \Delta t + q_t,$$  

where, in equilibrium,

$$q_t = \Delta x_t.$$

The market makers continue to maximize a quadratic objective:

$$\max_q \left\{ \hat{E}_t \sum_{s=t+n\Delta_t} e^{-r(s-t)} \left( \psi I_s^\top p_s \Delta_t q_s^\top \hat{p}_s - \frac{\gamma^M}{2} I_s^\top V_{\Delta_t} I_s \right) \right\},$$

as assumed in Section 3. One can therefore interpret $\Delta t$ in this case as the frequency with which the researcher observes the world, which does not impact (to the first order) equilibrium quantities — in particular, flow trades and costs.
subject to (A.6) and expectations about $q$ described below. Note that the market maker’s objective depends (positively) on the expected cash flows $\psi I_s^\top p_s \Delta_t - q_s^\top \hat{p}_s$ due to future trades with the end user and the trader and negatively on the risk of his inventory.

We assume that market makers cannot predict the trader’s order flow $\Delta x$. More specifically, according to their probability distribution,

$$\hat{E}_t[\Delta x_t | \mathcal{F}_s, s < t] = 0$$  \hspace{1cm} (A.8)

$$\hat{E}_t[(\Delta x_t)^2 | \mathcal{F}_s, s < t] = v.$$  \hspace{1cm} (A.9)

Moments of $q_s$ and $I_s$ follow immediately.

The first-order condition with respect to $q_t$ is

$$0 = \hat{E}_t \sum_{s = t + n\Delta_t} e^{-r(s-t)} \left( \psi p_s^\top - \gamma^M I_s^\top \frac{V_{\Delta_t}}{\Delta_t} \right) \frac{\partial I_s}{\partial q_t} \Delta_t - \hat{p}_t^\top.$$  \hspace{1cm} (A.10)

Using the fact that $\frac{\partial I_s}{\partial q_t} = (1 - \psi \Delta_t)^{s-t}$, the first-order condition yields

$$\hat{p}_t = \hat{E}_t \sum_{s = t + n\Delta_t} e^{-r(s-t)} e^{-r(s-t)} (1 - \psi \Delta_t)^{s-t} \left( \psi p_s - \gamma^M V_{\Delta_t} I_s \right) \Delta_t.$$  \hspace{1cm} (A.11)

Using the fact that $\hat{E}_t[e^{-r(s-t)} p_s] = p_t$, we obtain

$$\hat{p}_t = p_t - \kappa_I I_t$$  \hspace{1cm} (A.12)

for a constant matrix

$$\kappa_I = \sum_{n=0}^{\infty} e^{-r n \Delta_t} (1 - \psi \Delta_t)^{n \Delta_t} \gamma^M \frac{V_{\Delta_t}}{\Delta_t} \Delta_t.$$  \hspace{1cm} (A.13)

As $\Delta_t \to 0$, (A.12) continues to hold with

$$\kappa_I = \gamma^M (r + \psi)^{-1} \lim_{\Delta_t \to 0} \frac{V_{\Delta_t}}{\Delta_t}$$  \hspace{1cm} (A.14)

$$= \gamma^M (r + \psi)^{-1} \Sigma.$$  \hspace{1cm} (A.15)
Note that this price specification is the same as in Section 4, with

\[ D_t = -\kappa_I I_t \]  
(A.16)

\[ dD_t = -\kappa_I dI_t \]
\[ = -\kappa_I (-\psi I_t + \tau_t) dt \]
\[ = -\kappa_I \psi \kappa_I^{-1} D_t dt - \kappa_I \tau_t dt \]
\[ = -RD_t dt + C^\top \tau_t dt. \]  
(A.17)

### A.3 Transitory and Persistent Price-Impact Components

The two types of price impact can obtain simultaneously in this model, provided that two kinds of market makers coexist and interact in particular ways. Specifically, suppose that one group of market makers transact with the trader. After a period of length \( h \), these market makers clear their inventories with a second group of market makers, who specialize in locating end users and trading with them. As in Section A.2, these market makers deplete their inventories only gradually (at a constant rate), giving rise to a persistent impact. The trader must compensate both groups of market makers for the risk taken, resulting in the two price-impact components.

### A.4 Connection between Discrete and Continuous Time

The continuous-time model, and therefore solution, are readily seen to be the limit of their discrete-time analogues when parameters are chosen consistently, adjusted for the length of the time interval between successive trading opportunities. The adjustment takes the
following form.

\[
\begin{align*}
\hat{\Lambda}(\Delta_t) &= \Delta_t^{-1} \Lambda \\
\hat{\Sigma}(\Delta_t) &= \Sigma \\
\hat{\Omega}(\Delta_t) &= \Omega \\
\hat{B}(\Delta_t) &= B \\
\hat{\Phi}(\Delta_t) &= \Phi \\
\hat{\rho}(\Delta_t) &= \rho \\
\hat{\gamma}(\Delta_t) &= \gamma \\
\hat{R}(\Delta_t) &= R \\
\hat{C}(\Delta_t) &= C.
\end{align*}
\tag{A.18}
\]

We remark that the adjustment to the trading cost \( \Lambda \) in Equation (A.18) is as implied by (A.4) with constant \( h \): quadratic in \( \Delta x \) with coefficient of order \( \Delta_t^{-1} \). When trading costs are proportional to \( \Sigma \), the equation for \( \lambda \) simply follows from the previous analysis and \( \Lambda = \lambda \Sigma \). The other adjustments are immediate. For instance, Equations (A.19)–(A.20) simply state that the variance is proportional to time.

**B Proofs**

**Proof of Propositions 1 and 2.** We calculate the expected future value function as

\[
E_t[V(x_t, f_{t+1})] = -\frac{1}{2} x_t^T A_{xx} x_t + x_t^T A_{xf} (I - \Phi) f_t + \frac{1}{2} f_t^T (I - \Phi)^T A_{ff} (I - \Phi) f_t + \frac{1}{2} E_t(\varepsilon_{t+1}^T A_{ff} \varepsilon_{t+1}) + A_0.
\tag{B.1}
\]
The agent maximizes the quadratic objective \(-\frac{1}{2}x^\top J_t x_t + x_t^\top j_t + d_t\) with

\[
J_t = \gamma \Sigma + \Lambda + (1 - \rho)A_{xx}
\]
\[
j_t = (B + (1 - \rho)A_{xf}(I - \Phi))f_t + \Lambda x_{t-1}
\]  
\[
d_t = -\frac{1}{2}x_{t-1}^\top \Lambda x_{t-1} + (1 - \rho) \left( \frac{1}{2}f_t^\top (I - \Phi)^\top A_{ff}(I - \Phi)f_t + \frac{1}{2}E_t(\varepsilon_{t+1}^\top A_{ff}\varepsilon_{t+1} + A_0) \right).
\]

The maximum value is attained by

\[
x_t = J_t^{-1}j_t,
\]
which proves (10).

The maximum value is equal to \(V(x_{t-1}, f_t) = \frac{1}{2}j_t^\top J_t^{-1}j_t + d_t\) and combining this with (7) we obtain an equation that must hold for all \(x_{t-1}\) and \(f_t\). This implies the following restrictions on the coefficient matrices:

\[
-A_{xx} = \Lambda(\gamma \Sigma + \Lambda + (1 - \rho)A_{xx})^{-1}\Lambda - \Lambda
\]  
\[
A_{xf} = \Lambda(\gamma \Sigma + \Lambda + (1 - \rho)A_{xx})^{-1}(B + (1 - \rho)A_{xf}(I - \Phi))
\]  
\[
A_{ff} = (B + (1 - \rho)A_{xf}(I - \Phi))^\top (\gamma \Sigma + \Lambda + (1 - \rho)A_{xx})^{-1}(B + (1 - \rho)A_{xf}(I - \Phi))
\]  
\[
+(1 - \rho)(I - \Phi)^\top A_{ff}(I - \Phi).
\]

We next derive the coefficient matrices \(A_{xx}, A_{xf},\) and \(A_{ff}\) by solving these equations. For this, we first rewrite Equation (B.4) by letting \(Z = \Lambda^{-\frac{1}{2}}A_{xx}\Lambda^{-\frac{1}{2}}\) and \(M = \Lambda^{-\frac{1}{2}}\Sigma\Lambda^{-\frac{1}{2}}\), which yields

\[
Z = I - (\gamma M + I + (1 - \rho)Z)^{-1},
\]
which is a quadratic with an explicit solution. Since all solutions \(Z\) can written as a limit of
polynomials of $M$, $Z$ and $M$ commute and the quadratic can be sequentially rewritten as

$$(1 - \rho)Z^2 + Z(I + \gamma M - (1 - \rho)I) = \gamma M$$

$$(Z + \frac{1}{2(1 - \rho)}(\rho I + \gamma M))^2 = \frac{\gamma}{1 - \rho} M + \frac{1}{4(1 - \rho)^2} (\rho I + \gamma M)^2,$$

resulting in

$$Z = \left( \frac{\gamma}{1 - \rho} M + \frac{1}{4(1 - \rho)^2} (\rho I + \gamma M)^2 \right)^{\frac{1}{2}} - \frac{1}{2(1 - \rho)} (\rho I + \gamma M) \quad \text{(B.7)}$$

$$A_{xx} = \Lambda^{\frac{1}{2}} \left[ \left( \frac{\gamma}{1 - \rho} M + \frac{1}{4(1 - \rho)^2} (\rho I + \gamma M)^2 \right)^{\frac{1}{2}} - \frac{1}{2(1 - \rho)} (\rho I + \gamma M) \right] \Lambda^{\frac{1}{2}}, \quad \text{(B.8)}$$

that is,

$$A_{xx} = \left( \frac{\gamma}{1 - \rho} \Lambda^{\frac{1}{2}} \Sigma \Lambda^{\frac{1}{2}} + \frac{1}{4(1 - \rho)^2} (\rho^2 \Lambda^2 + 2\rho \gamma \Lambda^{\frac{1}{2}} \Sigma \Lambda^{\frac{1}{2}} + \gamma^2 \Lambda^{\frac{1}{2}} \Sigma \Lambda^{-1} \Sigma \Lambda^{\frac{1}{2}}) \right)^{\frac{1}{2}}$$

$$- \frac{1}{2(1 - \rho)} (\rho \Lambda + \gamma \Sigma). \quad \text{(B.9)}$$

Note that the positive definite choice of solution $Z$ is the only one that results in a positive definite matrix $A_{xx}$.

In the case $\Lambda = \lambda \Sigma$ for some scalar $\lambda > 0$, the solution is $A_{xx} = a \Sigma$, where $a$ solves

$$-a = \frac{\lambda^2}{\gamma + \lambda + (1 - \rho)a} - \lambda \quad \text{(B.10)}$$

or

$$(1 - \rho)a^2 + (\gamma + \lambda \rho)a - \lambda \gamma = 0, \quad \text{(B.11)}$$

with solution

$$a = \frac{\sqrt{(\gamma + \lambda \rho)^2 + 4\gamma \lambda (1 - \rho)} - (\gamma + \lambda \rho)}{2(1 - \rho)}. \quad \text{(B.12)}$$

The other value-function coefficient determining optimal trading is $A_{xf}$, which solves the
linear equation (B.5). To write the solution explicitly, we note first that, from (B.4),

$$\Lambda(\gamma \Sigma + \Lambda + (1 - \rho)A_{xx})^{-1} = I - A_{xx}\Lambda^{-1}. \tag{B.13}$$

Using the general rule that $\text{vec}(XYZ) = (Z^\top \otimes X)\text{vec}(Y)$, we re-write (B.5) in vectorized form:

$$\text{vec}(A_{xf}) = \text{vec}((I - A_{xx}\Lambda^{-1})B) + ((1 - \rho)(I - \Phi)^\top \otimes (I - A_{xx}\Lambda^{-1}))\text{vec}(A_{xf}), \tag{B.14}$$

so that

$$\text{vec}(A_{xf}) = (I - (1 - \rho)(I - \Phi)^\top \otimes (I - A_{xx}\Lambda^{-1}))^{-1} \text{vec}((I - A_{xx}\Lambda^{-1})B). \tag{B.15}$$

In the case $\Lambda = \lambda \Sigma$, the solution is

$$A_{xf} = \lambda B((\gamma + \lambda + (1 - \rho)a)I - \lambda(1 - \rho)(I - \Phi))^{-1}$$

$$= \lambda B((\gamma + \lambda \rho + (1 - \rho)a)I + \lambda(1 - \rho)\Phi))^{-1} \tag{B.16}$$

$$= B \left(\frac{\gamma}{a} + (1 - \rho)\Phi\right)^{-1}. \tag{B.17}$$

Finally, $A_{ff}$ is calculated from the linear equation (B.6), which is of the form

$$A_{ff} = Q + (1 - \rho)(I - \Phi)^\top A_{ff}(I - \Phi) \tag{B.18}$$

with

$$Q = (B + (1 - \rho)A_{xf}(I - \Phi))^\top (\gamma \Sigma + \Lambda + (1 - \rho)A_{xx})^{-1}(B + (1 - \rho)A_{xf}(I - \Phi))$$

a positive-definite matrix.

The solution is easiest to write explicitly for diagonal $\Phi$, in which case

$$A_{ff,ij} = \frac{Q_{ij}}{1 - (1 - \rho)(1 - \Phi_{ii})(1 - \Phi_{jj})}. \tag{B.19}$$
In general,

\[
\text{vec}(A_{ff}) = (I - (1 - \rho)(I - \Phi)^\top \otimes (I - \Phi)^\top)^{-1}\text{vec}(Q). \tag{B.20}
\]

One way to see that \(A_{ff}\) is positive definite is to iterate (B.18) starting with \(A_{ff}^0 = 0\), given that \(I \geq I - \Phi\).

Having computed the coefficient matrices, finishing the proof is straightforward. Equation (16) follows directly from (10). Equation (12) follows from (16) by using the equations for \(A_{xf}\) and \(a\), namely (B.5) and (B.10). \(\blacksquare\)

**Proof of Propositions 3 and 4.** Given the conjectured value function, the optimal choice \(\tau\) equals

\[
\tau_t = -\Lambda^{-1}A_{xx}x_t + \Lambda^{-1}A_{xf}f_t,
\]

Once this is inserted in the HJB equation, it results in the following equations defining the value-function coefficients (using the symmetry of \(A_{xx}\)):

\[
\begin{align*}
-\rho A_{xx} & = A_{xx}\Lambda^{-1}A_{xx} - \gamma \Sigma \tag{B.21} \\
\rho A_{xf} & = -A_{xx}\Lambda^{-1}A_{xf} - A_{xf}\Phi + B \tag{B.22} \\
\rho A_{ff} & = A_{xf}^\top\Lambda^{-1}A_{xf} - 2A_{ff}\Phi. \tag{B.23}
\end{align*}
\]

Pre- and post-multiplying (B.21) by \(\Lambda^{-\frac{1}{2}}\), we obtain

\[
-\rho Z = Z^2 + \frac{\rho^2}{4}I - C, \tag{B.24}
\]

that is,

\[
\left(Z + \frac{\rho}{2}I\right)^2 = C, \tag{B.25}
\]
where

\[ Z = \Lambda^{-\frac{1}{2}} A_{xx} \Lambda^{-\frac{1}{2}} \quad \text{(B.26)} \]

\[ C = \gamma \Lambda^{-\frac{1}{2}} \Sigma \Lambda^{-\frac{1}{2}} + \frac{\rho^2}{4} I. \quad \text{(B.27)} \]

This leads to

\[ Z = -\frac{\rho}{2} I + C^{\frac{1}{2}} \geq 0, \quad \text{(B.28)} \]

implying that

\[ A_{xx} = -\frac{\rho}{2} \Lambda + \Lambda^{\frac{1}{2}} \left( \gamma \Lambda^{-\frac{1}{2}} \Sigma \Lambda^{-\frac{1}{2}} + \frac{\rho^2}{4} \right)^{\frac{1}{2}} \Lambda^{\frac{1}{2}}. \quad \text{(B.29)} \]

The solution for \( A_{xf} \) follows from Equation (B.22), using the general rule that \( \text{vec}(XYZ) = (Z^\top \otimes X) \text{vec}(Y) \):

\[ \text{vec}(A_{xf}) = \left( \rho I + \Phi^\top \otimes I_K + I_S \otimes (A_{xx} \Lambda^{-1}) \right)^{-1} \text{vec}(B) \]

If \( \Lambda = \lambda \Sigma \), then \( A_{xx} = a \Sigma \) with

\[ -\rho a = a^2 \frac{1}{\lambda} - \gamma \quad \text{(B.30)} \]

with solution

\[ a = -\frac{\rho}{2} \lambda + \sqrt{\gamma \lambda + \frac{\rho^2}{4} \lambda^2}. \quad \text{(B.31)} \]

In this case, (B.22) yields

\[ A_{xf} = B \left( \rho I + \frac{a}{\lambda} I + \Phi \right)^{-1} \]

\[ = B \left( \frac{\gamma}{a} I + \Phi \right)^{-1}, \]

where the last equality uses (B.30).
Then, we have
\[
\tau_t = \frac{a}{\lambda} \left[ \Sigma^{-1} B \left( a\Phi + \gamma I \right)^{-1} f_t - x_t \right].
\] (B.32)

It is clear from (B.31) that \( \frac{a}{\lambda} \) decreases in \( \lambda \) and increases in \( \gamma \). ■

**Proof of Proposition 5.** We start with the continuous-time case. With

\[
\Pi = \begin{bmatrix} \Phi & 0 \\ 0 & R \end{bmatrix} \tag{B.33}
\]

\[
\tilde{C} = \begin{bmatrix} 0 \\ C \end{bmatrix} \tag{B.34}
\]

\[
\tilde{B} = \begin{bmatrix} B & -R \end{bmatrix} \tag{B.35}
\]

\[
\tilde{\Omega} = \begin{bmatrix} \Omega & 0 \\ 0 & 0 \end{bmatrix}, \tag{B.36}
\]

the HJB equation is

\[
\rho V = \max_{\tau} \left\{ x^\top \left( \tilde{B} y + C \tau \right) - \frac{\gamma}{2} x^\top \Sigma x - \frac{1}{2} \tau^\top \Lambda \tau + \frac{\partial V}{\partial x} \tau + \frac{\partial V}{\partial y} \left( -\Pi y + \tilde{C} \tau \right) + e \right\}
\]

\[
= \max_{\tau} \left\{ x^\top \tilde{B} y - \frac{\gamma}{2} x^\top \Sigma x - \frac{1}{2} \tau^\top \Lambda \tau + \tau^\top \left( Q_x x + Q_y y \right) - \frac{\partial V}{\partial y} \Pi y + e \right\},
\]

where

\[
e = \frac{1}{2} tr \left( \tilde{\Omega} \frac{\partial^2 V}{\partial y \partial y^\top} \right) \tag{B.37}
\]

\[
Q_x = -A_{xx} + \tilde{C}^\top A_{xy} + C^\top
\]

\[
Q_y = A_{xy} + \tilde{C}^\top A_{yy}. \tag{B.38}
\]
It follows immediately that

\[
\tau = -\Lambda^{-1}Q_x [\text{aim} - x] \tag{B.39}
\]

\[
= \Lambda^{-1} \left( A_{xx} - \tilde{C}^\top A_{xy} - C^\top \right) [\text{aim} - x]
\]

\[
\equiv \bar{M}^{\text{speed}} [\text{aim} - x],
\]

with

\[
\text{aim} = - (Q_y^{-1}Q_y) y
\]

\[
= \left( A_{xx} - \tilde{C}^\top A_{xy} - C^\top \right)^{-1} \left( A_{xy} + \tilde{C}^\top A_{yy} \right) y
\]

\[
\equiv \bar{M}^{\text{aim}} y.
\]

The coefficient matrices solve the system

\[
-\rho A_{xx} = -\gamma \Sigma + Q_x^\top \Lambda^{-1} Q_x
\]

\[
= -\gamma \Sigma + \left( A_{xx} - A_{xy} \tilde{C} - C \right) \Lambda^{-1} \left( A_{xx} - \tilde{C}^\top A_{xy} - C^\top \right)
\]

\[
\rho A_{xy} = Q_x^\top \Lambda^{-1} Q_y + \tilde{B} - A_{xy} \Pi
\]

\[
= - \left( A_{xx} - A_{xy} \tilde{C} - C \right) \Lambda^{-1} \left( A_{xy} + \tilde{C}^\top A_{yy} \right) + \tilde{B} - A_{xy} \Pi
\]

\[
\rho A_{yy} = Q_y^\top \Lambda^{-1} Q_y - 2A_{yy} \Pi
\]

\[
= \left( A_{xy}^\top + A_{yy} \tilde{C} \right) \Lambda^{-1} \left( A_{xy} + \tilde{C}^\top A_{yy} \right) - 2A_{yy} \Pi.
\]

We note that the equations above have to be solved simultaneously for \( A_{xx}, A_{xy}, \) and \( A_{yy} \); there is no closed-form solution. The complication is due to the fact that current trading affects the persistent price component \( D \) (that is, \( C \neq 0 \)).

We turn now to the discrete-time case. Throughout this discussion, \( y_t = (f_t^\top, D_{t-1}^\top)^\top \) (a column vector). Given this definition, it follows that

\[
E_t [y_{t+1} | x_t] = (I - \Pi) y_t + \tilde{C}(x_t - x_{t-1}), \tag{B.42}
\]

with \( \Pi \) and \( \tilde{C} \) as defined above.
The conjectured value function is

\[ V(x_{t-1}, y_t) = -\frac{1}{2} x_{t-1}^\top A_{xx} x_{t-1} + x_{t-1}^\top A_{xy} y_t + \frac{1}{2} y_t^\top A_{yy} y_t, \]  

(B.43)

so that

\[
E_t \left[ V(x_t, y_{t+1}) \mid x_t \right] = -\frac{1}{2} x_t^\top A_{xx} x_t + x_t^\top A_{xy} \left( (I - \Pi) y_t + \tilde{C}(x_t - x_{t-1}) \right) + \\
\frac{1}{2} \left( (I - \Pi) y_t + \tilde{C}(x_t - x_{t-1}) \right)^\top A_{yy} \left( (I - \Pi) y_t + \tilde{C}(x_t - x_{t-1}) \right) + \\
\frac{1}{2} E_t \left[ \tilde{\varepsilon}_{t+1}^\top A_{yy} \tilde{\varepsilon}_{t+1} \right]. 
\]

(B.44)

The trader consequently chooses \( x_t \) to solve

\[
\max_x \left\{ x^\top \left( \tilde{B} y_t + C(x - x_{t-1}) \right) - \frac{\gamma}{2} x^\top \Sigma x - \frac{1}{2} (x - x_{t-1})^\top \Lambda (x - x_{t-1}) + \\
(1 - \rho) \left[ -\frac{1}{2} x^\top A_{xx} x + x^\top A_{xy} \left( (I - \Pi) y_t + \tilde{C}(x - x_{t-1}) \right) + \\
\frac{1}{2} \left( (I - \Pi) y_t + \tilde{C}(x - x_{t-1}) \right)^\top A_{yy} \left( (I - \Pi) y_t + \tilde{C}(x - x_{t-1}) \right) \right] \right\}, 
\]

(B.45)

which is a quadratic of the form \(-\frac{1}{2} x^\top J x + x^\top j_t + d_t\), with

\[
J = \gamma \Sigma + \Lambda + (1 - \rho) A_{xx} - 2 C - 2(1 - \rho) A_{xy} \tilde{C} - (1 - \rho) \tilde{C}^\top A_{yy} \tilde{C} 
\]

(B.46)

\[
\dot{j}_t = \tilde{B} y_t - C x_{t-1} + \Lambda x_{t-1} + (1 - \rho) A_{xy} \left( (I - \Pi) y_t + \tilde{C} x_{t-1} \right) + \\
(1 - \rho) \tilde{C} A_{yy} \left( (I - \Pi) y_t - \tilde{C} x_{t-1} \right) 
\]

(B.47)

\[
\equiv S_x x_{t-1} + S_y y_t 
\]

(B.48)

\[
d_t = -\frac{1}{2} x_{t-1}^\top \Lambda x_{t-1} - \frac{1 - \rho}{2} \left( (I - \Pi) y_t + \tilde{C} x_{t-1} \right)^\top A_{yy} \left( (I - \Pi) y_t - \tilde{C} x_{t-1} \right). 
\]

(B.49)

Here,

\[
S_x = \Lambda - C - (1 - \rho) A_{xy} \tilde{C} - (1 - \rho) \tilde{C}^\top A_{yy} \tilde{C} 
\]

(B.50)

\[
S_y = \tilde{B} + (1 - \rho) A_{xy} (I - \Pi) - (1 - \rho) \tilde{C}^\top A_{yy} (I - \Pi). 
\]

(B.51)
The value of \( x \) attaining the maximum is given by

\[
x_t = J^{-1} j_t,
\]
and the maximal value is

\[
\frac{1}{2} j_t J^{-1} j_t + d_t = V(x_{t-1}, y_t) \tag{B.52}
\]

\[
= -\frac{1}{2} x_{t-1}^\top A_{xx} x_{t-1} + x_{t-1}^\top A_{xy} y_t + \frac{1}{2} y_t^\top A_{yy} y_t. \tag{B.53}
\]

The unknown matrices have to satisfy a system of equations encoding the equality of all coefficients in (B.54). Thus,

\[
-A_{xx} = S_x^\top J^{-1} S_x - \Lambda - (1 - \rho) \hat{C}^\top A_{yy} \hat{C} \tag{B.55}
\]

\[
A_{xy} = S_x^\top J^{-1} S_y + (1 - \rho) \hat{C}^\top A_{yy} (I - \Pi) \tag{B.56}
\]

\[
A_{yy} = S_y^\top J^{-1} S_y + (1 - \rho) (I - \Pi)^\top A_{yy} (I - \Pi). \tag{B.57}
\]

For our purposes, the more interesting observation is that the optimal position \( x_t \) is rewritten as

\[
x_t = x_{t-1} + \left( I - J^{-1} S_x \right) \left( \left( I - J^{-1} S_x \right)^{-1} (JS_y y_t) - x_{t-1} \right). \tag{B.58}
\]

Using (B.55), the “speed” can also be expressed as

\[
M^{speed} \equiv I - J^{-1} S_x
\]

\[
= (S_x^\top)^{-1} (S_x^\top - S_x J^{-1} S_x)
\]

\[
= (S_x^\top)^{-1} (S_x^\top + A_{xx} - \Lambda + (1 - \rho) \hat{C} A_{yy} \hat{C}^\top)
\]

\[
= (S_x^\top)^{-1} \left( A_{xx} - C^\top - (1 - \rho) \hat{C} A_{xy}^\top \right). \tag{B.59}
\]
while, using (B.56), the aim is

$$M^{aim} \equiv (I - J^{-1}S_x)^{-1}JS_y y_t$$
$$= (I - J^{-1}S_x)^{-1}(S_x^\top)^{-1}S_x^\top JS_y y_t$$
$$= (I - J^{-1}S_x)^{-1}(S_x^\top)^{-1}\left(A_{xy} + (1 - \rho)\tilde{C}^\top A_{yy}(I - \Pi)\right)$$
$$= \left(A_{xx} - C - (1 - \rho)\tilde{C}A_{xy}^\top\right)^{-1}\left(A_{xy} + (1 - \rho)\tilde{C}^\top A_{yy}(I - \Pi)\right). \quad (B.60)$$

**Proof of Proposition 6.** We work with the characterization of solutions provided in the proof of Proposition 5. We show that, as $\Delta_t \to 0$, Equations (B.55)–(B.57) tend to their counterparts in (B.41), which implies that the solutions also do.

To keep the proof short, we prove the claim only for (B.55); the other two equations work similarly. We first rewrite this equation as

$$-A_{xx} = S_x^\top J^{-1}S_x - \Lambda - (1 - \hat{\rho}(\Delta_t))\tilde{C}^\top A_{yy}\tilde{C}$$
$$= (S_x - J)^\top J^{-1}(S_x - J) - \Lambda - J + 2S - (1 - \hat{\rho}(\Delta_t))\tilde{C}^\top A_{yy}\tilde{C}, \quad (B.61)$$

and then rearrange it as

$$-A_{xx} \left(1 - e^{-\rho \Delta_t}\right) = (S_x - J)^\top J^{-1}(S_x - J) - \gamma \Sigma \Delta_t. \quad (B.62)$$

Dividing through by $\Delta_t$ and ignoring terms in $\Delta_t$ in $S_x - J$ and $J\Delta_t$, we obtain

$$-\rho A_{xx} = -\gamma \Sigma + \left(A_{xx} - A_{xy}\tilde{C} - C\right)^\top \Lambda^{-1} \left(A_{xx} - \tilde{C}^\top A_{xy} - C^\top\right), \quad (B.63)$$

the same as in continuous time.

Having established that the value function coefficients in discrete time have as limit their counterparts in continuous time, we now note that, when letting $\Delta_t$ go to 0, the speed term $M^{speed}$ is given by

$$\Lambda^{-1} \left(A_{xx} - C - \tilde{C}A_{xy}^\top\right), \quad (B.64)$$
while the aim term $M^{\text{aim}}$ by

$$
(A_{xx} - C^T A_{xy}^T - C^T)^{-1} \left( A_{xy} + \tilde{C}^T A_{yy} \right) .
$$

(B.66)

These expressions are the same as obtained in continuous time. □

**Proof of Proposition 7.** Suppose that $\alpha_t = B f_t$ with $B$ given by (53) and apply Proposition 4 to conclude that, if $x_t = -f_t^{K+1}$, then $dx_t = -df_t^{K+1}$. The comparative-static results are immediate. □
References


Lillo, F., J. D. Farmer, and R. N. Mantegna, 2003, “Master Curve for Price-Impact Func-


press.


Lynch, A., and P. Balduzzi, 2000, “Predictability and transaction costs: the impact on

Return Predictability, Wealth Shocks and State-dependent Transaction Costs.,” Working
paper, NYU.

Obizhaeva, A., and J. Wang, 2006, “Optimal trading strategy and supply/demand dynam-
ics,” Working paper, MIT.


Management, 14, 4–9.


Vayanos, D., and J.-L. Vila, 1999, “Equilibrium Interest Rate and Liquidity Premium with
<table>
<thead>
<tr>
<th>Commodity</th>
<th>Average Price Per Contract</th>
<th>Standard Deviation of Price Changes</th>
<th>Contract Multiplier</th>
<th>Daily Trading Volume (Contracts)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aluminum</td>
<td>44,561</td>
<td>637</td>
<td>25</td>
<td>9,160</td>
</tr>
<tr>
<td>Cocoa</td>
<td>15,212</td>
<td>313</td>
<td>10</td>
<td>5,320</td>
</tr>
<tr>
<td>Coffee</td>
<td>38,600</td>
<td>1,119</td>
<td>37,500</td>
<td>5,640</td>
</tr>
<tr>
<td>Copper</td>
<td>80,131</td>
<td>2,023</td>
<td>25</td>
<td>12,300</td>
</tr>
<tr>
<td>Crude</td>
<td>40,490</td>
<td>1,103</td>
<td>1,000</td>
<td>151,160</td>
</tr>
<tr>
<td>Gasoil</td>
<td>34,963</td>
<td>852</td>
<td>100</td>
<td>37,260</td>
</tr>
<tr>
<td>Gold</td>
<td>43,146</td>
<td>621</td>
<td>100</td>
<td>98,700</td>
</tr>
<tr>
<td>Lead</td>
<td>23,381</td>
<td>748</td>
<td>25</td>
<td>2,520</td>
</tr>
<tr>
<td>Natgas</td>
<td>50,662</td>
<td>1,932</td>
<td>10,000</td>
<td>46,120</td>
</tr>
<tr>
<td>Nickel</td>
<td>76,530</td>
<td>2,525</td>
<td>6</td>
<td>1,940</td>
</tr>
<tr>
<td>Silver</td>
<td>36,291</td>
<td>893</td>
<td>5,000</td>
<td>43,780</td>
</tr>
<tr>
<td>Sugar</td>
<td>10,494</td>
<td>208</td>
<td>112,000</td>
<td>25,700</td>
</tr>
<tr>
<td>Tin</td>
<td>38,259</td>
<td>903</td>
<td>5</td>
<td>NaN</td>
</tr>
<tr>
<td>Unleaded</td>
<td>47,967</td>
<td>1,340</td>
<td>42,000</td>
<td>11,320</td>
</tr>
<tr>
<td>Zinc</td>
<td>36,513</td>
<td>964</td>
<td>25</td>
<td>6,200</td>
</tr>
</tbody>
</table>

Table 1: **Summary Statistics.** For each commodity used in our empirical study, the first column reports the average price per contract in U.S. dollars over our sample period 01/01/1996–01/23/2009. For instance, since the average gold price is $431.46 per ounce, the average price per contract is $43,146 since each contract is for 100 ounces. Each contract’s multiplier (100 in the case of gold) is reported in the third column. The second column reports the standard deviation of price changes. The fourth column reports the average daily trading volume per contract, estimated as the average daily volume of the most liquid contract traded electronically and outright (i.e., not including calendar-spread trades) in December 2010.
Table 2: Performance of Trading Strategies Before and After Transaction Costs.
This table shows the annualized Sharpe ratio gross and net of trading costs for the optimal trading strategy in the absence of trading costs (“no TC”), our optimal dynamic strategy (“optimal”), and a strategy that optimizes a static one-period problem with trading costs (“static”). Panel A illustrates this for a low transaction cost parameter, while Panel B has a high one.
Construction of Current Optimal Trade

Figure 1: Optimal Trading Strategy: Triangulating between Current, Markowitz, and Future Aim Portfolios. This figure shows how the optimal trade moves the portfolio from the existing position $x_{t-1}$ towards the aim, trading only part of the way to the aim to limit transactions costs. The aim is an average of the static Markowitz portfolio and the expected future aim, which incorporates the expected alpha decay.
Figure 2: Optimal Trading Strategy: Chase Aim Underweighting Fast-Decay Factors. This figure shows how the optimal trade moves from the existing position $x_{t-1}$ towards the aim, which puts relatively more weight on assets loading on persistent factors. Relative to the Markowitz portfolio, the weight in the aim of asset 2 is lower because asset 2 has a faster-decaying alpha, as is apparent in the lower expected weight it receives in future Markowitz portfolios.
Figure 3: **Positions in Crude and Gold Futures.** This figure shows the positions in crude and gold for the optimal trading strategy in the absence of trading costs ("Markowitz") and our optimal dynamic strategy ("optimal").
Figure 4: Optimal Trading in Response to Shock to Return Predicting Signals. This figure shows the response in the optimal position following a shock to a return predictor as a function of the number of days since the shock. The top left panel does this for a shock to the fast 5-day return predictor, the top right panel considers a shock to the 12-month return predictor, and the bottom panel to the 5-year predictor. In each case, we consider the response of the optimal trading strategy in the absence of trading costs (“Markowitz”) and our optimal dynamic strategy (“optimal”) using high and low transactions costs.