OPTIMAL TRANSACTION FILTERS UNDER TRANSITORY TRADING OPPORTUNITIES: Theory and Empirical Illustration*

Ronald Balvers and Yangru Wu

Working Paper Series WCRFS: 08-21
If transitory profitable trading opportunities exist, filter rules mitigate transaction costs. We use a dynamic programming framework to design an optimal filter which maximizes after-cost expected returns. The filter size depends crucially on the degree of persistence of trading opportunities, transaction cost, and standard deviation of shocks. Applying our theory to daily dollar-yen exchange trading, we find that the optimal filter can be economically significantly different from a naïve filter equal to the transaction cost. The candidate trading strategies generate positive returns that disappear after accounting for transaction costs. However, when the optimal filter is used, returns after costs remain positive and are higher than for naïve filters.

Keywords: Transaction Costs, Filter Rules, Trading Strategies, Foreign Exchange
JEL Codes: G10, G15, G11
OPTIMAL TRANSACTION FILTERS UNDER TRANSITORY TRADING OPPORTUNITIES:
Theory and Empirical Illustration

April 30, 2008

ABSTRACT
If transitory profitable trading opportunities exist, filter rules mitigate transaction costs. We use a dynamic programming framework to design an optimal filter which maximizes after-cost expected returns. The filter size depends crucially on the degree of persistence of trading opportunities, transaction cost, and standard deviation of shocks. Applying our theory to daily dollar-yen exchange trading, we find that the optimal filter can be economically significantly different from a naïve filter equal to the transaction cost. The candidate trading strategies generate positive returns that disappear after accounting for transaction costs. However, when the optimal filter is used, returns after costs remain positive and are higher than for naïve filters.

Keywords: Transaction Costs, Filter Rules, Trading Strategies, Foreign Exchange
JEL Codes: G10, G15, G11

Introduction
It is inarguable that opportunities for above-normal returns are available to market participants at some level. These opportunities may be exploitable for instance at an intra-daily frequency as a reward for information acquisition when markets are efficient, or at a lower frequency to market timers when markets are inefficient. By nature these profit opportunities are predicable but transitory, and transaction costs may be a major impediment in exploiting them. This paper explores the optimal trading strategy when transitory opportunities exist and transactions are costly.

The model we present is applicable to the arbitraging of microstructure inefficiencies that require frequent and timely transactions, which may be largely riskless. An example is uncovered interest speculation in currency market where a trader takes either one side of the market or the reverse. Alternatively, a trader arbitrages differences between an asset’s return and that of one of its derivatives: going long on the arbitrage position or reversing the position and going short, as is the case in covered

\[1\] For instance, Grundy and Martin (2001) express doubt that the anomalous momentum profits survive transaction costs, and Hanna and Ready (2001) find that the momentum profits are substantially reduced when transactions costs are accounted for. Lesmond et al. (2004) conclude more strongly that momentum profits with transactions costs are illusory. Neely and Weller (2003) reach a similar conclusion for trading profits in foreign exchange markets after transaction costs.
interest arbitrage.

Starting with Alexander (1961) and Fama and Blume (1966), trading rules often involve a “filter” that prompts one to trade only if a realization exceeds a benchmark by a certain percentage. For instance in foreign exchange trading an x percent filter might stipulate that a currency be purchased if its current exchange value exceeds some moving average of its past values by more than the x percent “band”. Presumably, such a filter reduces trading and thus transaction costs. The idea is that if the trade indicator is weak the expected return from the transaction may not compensate for the transaction cost. Lehmann (1990) provides an interesting alternative filter by varying portfolio weights according to the strength of the return indicators – in trading smaller quantities of the assets with the weaker trade indicators, transaction costs are automatically reduced relative to the payoff.

Knez and Ready (1996) and Cooper (1999) explore different filters and find that the after-transaction-cost returns indeed improve compared to trading strategies with zero filter. The problem with the filter approach is that there is no way of knowing a priori which filter band would be reasonable because the buy/sell signal and the transaction cost are not in the same units – the filter is the percentage by which the effective signal exceeds the signal at which a change in position first appears profitable before transaction costs, but this percentage bears no relation to the percentage return expected. This also implies that there is no discipline against data mining for researchers: many filters with different bands can be tried to fabricate positive net strategy returns. While Lehmann’s (1990) approach provides more discipline as it specifies a unique strategy, the filter it implies is not necessarily optimal.

The purpose of this paper is to design an optimal filter that a priori maximizes the expected return net of transaction cost. To accomplish this we employ a “parametric” approach (see for instance Balvers, Wu, and Gilliland (2000)) that allows the trading signal and the transaction cost to be in the same units. In effect we convert a filter into returns space and then are able to derive the filter’s optimal band. The optimal filter depends on the exact balance between maintaining the most profitable transactions and minimizing the transactions costs.

The optimal filter (band) can be no larger than the transaction cost (plus interest). This is clear
because there is no reason to exclude trades that have an immediate expected return larger than the transaction cost. In general the optimal filter is significantly smaller than the transaction cost. This occurs when the expected return is persistent: even if the immediate return from switching is less than the transaction cost, the persistence of the expected return makes it likely that an additional return is foregone in future periods by not switching. Roughly, the filter must depend on the transaction cost as well as a factor related to the probability that a switch occurs. Our model characterizes the determinants of the filter in general and provides an exact solution for the filter when zero-investment returns have a uniform distribution.

In exploring the effect of transaction costs when returns are predictable, this paper has the same objective as Balduzzi and Lynch (1999), Lynch and Balduzzi (2000), and Lynch and Tan (2002). The focus of these authors, however, differs significantly from ours in that they consider the utility effects and portfolio rebalancing decisions, respectively, in a life cycle portfolio choice framework. They simulate the welfare cost and portfolio rebalancing decisions given a trader’s constant relative risk aversion utility function, but they do not provide analytical solutions and it is difficult to use their approach to quantify the optimal trading strategies for particular applications. Our approach, in contrast, provides specific theoretical results yielding insights into the factors affecting optimal trading strategies. Moreover our results can be applied based on observable market characteristics that do not depend on subjective utility function specifications.

In contrast to Balduzzi and Lynch, Lynch and Balduzzi, and Lynch and Tan, we sidestep the controversial issue of risk in the theory. This simplifies our analysis considerably and is reasonable in a variety of circumstances. First, we can think of the raw returns as systematic-risk-adjusted returns, with whichever risk model is considered appropriate. The systematic risk adjustment is sufficient to account for all risk as long as trading occurs at the margins of an otherwise well-diversified portfolio. Second, in

---

2 There is a far more extensive literature considering investment choices under transaction costs when returns are not predictable. See Liu and Loewenstein (2002) and Liu (2004) for recent examples. Marquering and Verbeek (2004) assume predictability and adjust for transactions costs. Their approach, however, is a complement to ours in that they integrate risk into the optimal switching choice while accounting for transactions costs after the fact, whereas our approach integrates transactions costs into the optimal switching choice while accounting for risk after the fact.
particular at intra-daily frequencies, traders may create arbitrage positions so that risk is irrelevant. Third, in many applications risk considerations are perceived as secondary compared to the gains in expected return; if risk adjustments are relatively small so that the optimal trading rules are approximately correct then risk corrections can be safely applied to *ex post* returns.

Our framework implies that trading strategy expected returns using the optimal filter should be smaller than the expected returns when transaction costs are ignored but larger than for any other filter strategy with transaction costs, in particular those for which the filter is zero percent or equal to the transaction cost. We apply our optimal filter to a natural case for our model: daily foreign exchange trading in the yen/dollar market. As is well-known (see for instance Cornell and Dietrich (1978), Sweeney (1986), LeBaron (1998), Gencay (1999), and Qi and Wu (2006)), simple moving-average trading rules improve forecasts of exchange rates and generate positive expected returns (with or without risk adjustment) in the foreign exchange market. However, for daily trading, returns net of transaction costs are negative or insignificant if no filter is applied (Neely and Weller (2003)).

We find that for the optimal filter the net returns are still significantly positive and higher than those when the filter is set equal to the transaction cost. Further, the optimal filter derived from the theory given a uniform distribution and two optimal filters derived numerically under normality and bootstrapping assumptions all generate similar results that are relatively close to the *ex post* maximizing filter for actual data. These results are important as they suggest an approach for employing trading strategies with filters to deal with transactions cost, without inviting data mining. The results also hint that in some cases the conclusion that abnormal profits disappear after accounting for transaction costs may be worth revisiting.

The next section develops the theoretical model and provides a general characterization of the optimal filter for an ARMA(1,1) returns process with general shocks, as well as a specific formula for the case when the shocks follow the uniform distribution. In section II, we apply the model to uncovered

---

3 Gencay, *et al.* (2002) show that real time trading models which employ more sophisticated techniques than the simple moving average rules can generate positive cost-adjusted returns with intra-day data.
currency speculation. We show first that the moving-average strategy popular in currency trading can be related to our ARMA(1,1) specification. We then use the first one-third of our sample to develop estimates of the returns process which we employ to calculate the optimal filter for an ARMA(1,1), an AR(1), and two representative MA returns processes. The optimal filter is obtained from the theoretical model for the uniform distribution but also numerically for the normal distribution and the bootstrapping distribution. Section III then conducts the out-of-sample test with the final two-thirds of our sample to compare mean returns from a switching strategy before and after transaction costs. The switching strategies are conducted under a variety of filters, including the optimal ones, for each of the ARMA(1,1), AR(1), and MA returns cases. Section IV concludes the paper.

I. The Theoretical Model

A. Autoregressive Conditional Returns and Two Risky Assets

An individual investor maximizes the discounted expected value of an investment over the infinite horizon. There is a proportional transaction cost and the investor chooses in each period between two assets that have autocorrelated mean returns. Each period, the investor is assumed to take a zero-cost investment position: a notional $1 long position in one asset and a notional $1 short position in the other asset. This implies that any profits or losses do not affect future investment positions. The return on the zero-cost investment position $x_t$ (a $1 long position in asset 1 and a $1 short position in asset 2) is assumed to follow an ARMA(1,1) process as a parsimonious parameterization of mild return predictability. Given that either the return on the investment position has no systematic risk or the investor is risk neutral, the decision problem is

$$V_1(x_{t-1}, \varepsilon_{t-1}) = E_{t-1} \left[ x_t + \max \left( \frac{V_1(x_t, \varepsilon_t)}{1+r}, \frac{V_2(x_t, \varepsilon_t)}{1+r} - 2c \right) \right],$$

subject to: $x_t = \rho x_{t-1} - \delta \varepsilon_{t-1} + \varepsilon_t$, $E_{t-1} \varepsilon_t = 0$, $\rho > \max(0, \delta)$ (2)

Each period, the investor chooses whether to hold a zero-investment position long in asset 1 and short in
asset 2, or the reverse. A proportional transaction cost $c$ is incurred whenever a position is closed out, implying a cost of $2c$ when a position is reversed.\footnote{Given the assumed risk neutrality, symmetry, and proportional transaction costs, intermediate positions, with investment in both assets or in neither asset, are never optimal.} In equation (1), the expected net present value of the investment strategy denoted by the value function $V_i(t)$ at time $t$ depends on the state as given by the existing long position in asset $i$ and short position in the other asset, and the variables describing the distribution of $x_t$, namely $x_{t-1}$ and $\varepsilon_{t-1}$. This value equals the current expected return, given the existing asset position (which we call 1 without loss of generality), plus the expected value in the next period discounted once at rate $r$, which depends on the updated return state variables $x_t$ and $\varepsilon_t$ as well as the new zero-investment asset position ($i = 1$ or 2, whichever maximizes the expected net present value), and minus the up-front adjustment cost incurred if the zero-investment position is switched from long in asset 1 to long in asset 2. Wealth is not a state variable, because of risk neutrality and the assumption that the long position in each period is always $1. Assuming that the scale of investment is low enough that wealth/margin does not become an issue preserves the relative simplicity of the decision problem.

Equation (2) describes the return process $x_t$ for the zero-investment position long in asset 1 and short in asset 2. The ARMA(1,1) process for $x_t$ is assumed to be persistent in the sense that $x_{t-1}$ positively affects $x_t$ ($\rho > 0$) and that $\varepsilon_{t-1}$ positively affects $x_t$ ($\rho > \delta$). The unconditional mean of $x_t$ is zero and reversing the zero-cost position necessarily generates a return of $-x_t$. Define $\mu_t \equiv E_{t-1} x_t$, and let it represent, without loss of generality, the conditional expected return of the current zero-investment position. The disturbance term $\varepsilon_t$ is symmetric and unimodal, with unbounded support, density $f(\varepsilon_t)$, and cumulative density $F(\varepsilon_t)$. $V(\cdot)$ denotes the maximum expected net present value of the strategy given the existing zero-investment position. Then:

**Proposition 1.** For the decision problem in equations (1) and (2) a unique $\mu^* < 0$ exists such that the investor maintains the current zero-investment position whenever
\( \mu_{t+1} > \mu^* \) and reverses the position whenever \( \mu_{t+1} \leq \mu^* \). The resulting expected net present value is:

\[
V(\mu_i) = \mu_i + \int_{\rho-\delta}^{\rho+\delta} \left( \frac{V(\mu_{t+1})}{1+r} \right) dF(\varepsilon_i) + \int_{-\infty}^{\rho-\delta} \left( \frac{V(-\mu_{t+1})}{1+r} - 2\varepsilon \right) dF(\varepsilon_i), \tag{3}
\]

subject to: \( \mu_{t+1} = \rho \mu_i + (\rho - \delta) \varepsilon_i \), with \( E_{t-1}\varepsilon_i = 0 \), \( \rho > \text{Max}(0, \delta) \). \tag{4}

**Proof.** See Appendix.

Equation (3) states that the maximum expected discounted investment return is equal to the expected return for the current position \( \mu_i \) plus, either the once-discounted maximum expected net present value of the strategy when the current investment position is maintained, or the once-discounted maximum expected net present value of the strategy, net of adjustment costs incurred when the current investment position is reversed. The integral bounds indicate the critical (cutoff) value for the current disturbance term \( \varepsilon_i \) according to which the investment position is maintained or reversed (higher \( \varepsilon_i \) implies higher future expected return for the existing zero-investment position so less incentive to switch).

\( \mu_i \) is a sufficient state variable for the individual investor maximizing the expected net present value from the zero-cost investment positions because equation (2) for the realized zero-investment return implies equation (4) for the conditional expected zero-investment return, and the latter is the pertinent variable for the risk-neutral expected value maximizer.

Given that the benefit of switching, the difference between \( V(-\mu) \) and \( V(\mu) \), is monotonically decreasing in \( \mu \) there is exactly one critical value \( \mu^* \) below which switching is optimal. Intuitively \( \mu^* \) must be negative because it makes no sense to switch when \( \mu_{t+1} \) is zero or positive. Our purpose is to provide in the following a specific characterization of \( \mu^* \) for empirical purposes.

To examine the specific advantage from switching the asset position, take the difference between the value of long in one asset and short in the other after a switch, \( V(-\mu) \), and the initial position, \( V(\mu) \):
\[ V(-\mu_t) - V(\mu_t) = -2\mu_t + B(\mu_t) + C(\mu_t) \]  

(5)

Equation (5) follows directly from applying equation (3) twice, for \(-\mu_t\) and \(\mu_t\), and manipulating the integrals (as shown in the Appendix). The first term on the right-hand side indicates the direct benefit of switching from \(\mu_t\) to \(-\mu_t\). The second term, \(B(\mu_t)\), is given as

\[ B(\mu_t) = \int_{\frac{-\mu_t - \rho\mu_t}{\rho - \delta}}^{\frac{-\mu_t + \rho\mu_t}{\rho - \delta}} \left( \frac{V(-\mu_{t+1}) - V(\mu_{t+1})}{1 + r} \right) dF(\varepsilon_i). \]  

(6)

\(B(\mu_t)\) represents the future benefit of switching now, given that neither of the two possible positions (the current available choices) would be switched in the next period. The integral bounds equal the critical values for \(\varepsilon_t\) at which a switch takes place (lower bound for switching next period when having switched in the previous period; upper bound for switching next period when not having switched previously). Use equation (4) to see that the difference \(V(-\mu_{t+1}) - V(\mu_{t+1})\) on the right-hand side of equation (6) is evaluated from \(\mu_t^*\) to \(-\mu_t^*\). The symmetry of this difference further implies that the sign depends only on the shape of the density function. In particular, taking \(\mu_t^* < 0\) (which is always so, as we show shortly), if \(\mu_t < 0\) then the range over which \(V(-\mu_{t+1}) - V(\mu_{t+1})\) is negative (whenever \(\mu_{t+1} > 0\)) is weighted less than the range over which it is positive, because the density is symmetric and unimodal (vice versa if \(\mu_t > 0\)). Thus \(B(\mu_t)\) is always strictly positive for \(\mu_t < 0\), unless the density function is flat (as is the case for the uniform distribution). The third term in equation (5), \(C(\mu_t)\), is given as

\[ C(\mu_t) = 2c \int_{\frac{-\mu_t + \rho\mu_t}{\rho - \delta}}^{\frac{-\mu_t - \rho\mu_t}{\rho - \delta}} dF(\varepsilon_i) = 2c \left[ F(\frac{-\mu_t - \rho\mu_t}{\rho - \delta}) - F(\frac{-\mu_t + \rho\mu_t}{\rho - \delta}) \right]. \]  

(7)

It gives the future benefit of switching now, given that for one of the two possible positions it will be optimal to switch in the next period (note that it is never optimal for both positions to switch). In that case, both possible positions end up becoming identical one period later and the only difference is the
transaction cost. \( C(\mu_i) \) represents the expected difference in transaction cost expenditure for the two positions: the cost times the probability of not switching initially but switching anyway next period minus the probability of switching initially and then switching back in the next period. Note that, from inspection of equation (7), \( C(\mu_i) \) must be positive for \( \mu_i < 0 \) (and negative for \( \mu_i > 0 \)).

To determine the critical value \( \mu^* \) we evaluate equation (5) for \( \mu_i = \mu^* \). First consider that \( \mu^* \) is determined optimally. Since the integral bounds are chosen optimally to obtain the maximum expected net present value from the investment strategy, it must be that the derivative with respect to the bound in equation (3) is zero. Using Leibniz’s rule to obtain the first-order condition for \( \mu^* \) in equation (3) gives

\[
\frac{V(-\mu^*)}{1+r} - \frac{V(\mu^*)}{1+r} = 2c
\]

Equation (8) reveals that the critical return is determined at the point where the total present value of expected benefit from reversing the investment position is exactly equal to the upfront transaction cost: at \( \mu_{\cdot,\cdot} = \mu^* < 0 \) the investor is indifferent between maintaining the current investment position with negative expected return \( \mu^* \) next period and reversing the investment position which has an immediate transaction cost of \( 2c \) but a positive expected next-period return \( -\mu^* \). It is directly clear from equation (8) that \( \mu^* \) must be constant over time as its notation without time subscript presumes.

Evaluate equation (5) at \( \mu_i = \mu^* \) and use equation (8) to obtain:

\[
\mu^* = -c(1+r) + \frac{1}{2}[B(\mu^*) + C(\mu^*)]
\]

Intuitively, equation (9) sets the critical expected return \( \mu^* \) equal to the up-front transaction cost plus interest saved by not switching, minus the cost of staying put at \( \mu^* \) one more period, which makes next-
period switching more likely, $C(\mu^*)$, or, if no switching, next period causes further low returns on average, $B(\mu^*)$. As $B(\mu^*) + C(\mu^*) > 0$ from equations (6) and (7), given that $\mu^* < 0$, it follows that $-\mu^* < c(1+r)$. This is intuitive because, whenever the current expected return is greater than or equal to the transaction cost, a switch must be optimal because the end-of-period gain in expected return $-2\mu^*$ immediately pays for the transaction cost $2c(1+r)$ and the switch also improves the position for future profits.

B. Closed Form Solution for Uniform Innovations

It is difficult to obtain an explicit analytical solution for the optimal filter in equation (9) because, from equation (6), $B(\mu^*)$ depends on the value function which is of unknown functional form. However, for the special case of a constant density over the relevant range (a uniform distribution), the $B(\mu^*)$ term simplifies substantially, as shown in the following, so that an explicit solution can be obtained.

Assume a uniform distribution for $\varepsilon$ over the interval $[-z, z]$ with implied density $f(\varepsilon)=1/(2z)$. To apply to this case, Proposition 1 requires minor modifications, which we omit for brevity, since the uniform distribution is bounded and not strictly unimodal. Equation (6) evaluated at $\mu^*$ becomes:

$$B(\mu^*) = \frac{1}{2z(1+r)} \int_{\operatorname{Max}[-z, \mu^*(1-\rho)/\rho-\delta]}^{\operatorname{Min}[z, -\mu^*(1+\rho)/\rho-\delta]} (V[-\rho\mu_i - (\rho-\delta)\varepsilon_i] - V[\rho\mu_i + (\rho-\delta)\varepsilon_i]) d\varepsilon_i. \quad (10)$$

Equation (7) evaluated at $\mu^*$ becomes:

$$C(\mu^*) = \frac{c}{z} \int_{\operatorname{Max}[-z, -\mu^*(1+\rho)/\rho-\delta]}^{\operatorname{Min}[z, \mu^*(1-\rho)/\rho-\delta]} d\varepsilon_i = \frac{c}{z} \left[ \min(z, \frac{\mu^*(1-\rho)}{\rho-\delta}) - \max(-z, \frac{\mu^*(1+\rho)}{\rho-\delta}) \right]. \quad (11)$$

\[\text{However, } C(\mu^*) = 0 \text{ when } \mu^*(1-\rho) < -(\rho-\delta)z. \text{ This condition and the Min and Max operators in equations (10) and (11) appear because } dF(\varepsilon) = 0 \text{ outside of the domain } [-z, z] \text{ but } d\varepsilon \text{ is not. In the following we will provide a condition to avoid these additional cases.}\]
Proposition 2. If \( \varepsilon \) is uniformly distributed over the interval \([-z, z]\) and if
\[
z \geq \frac{[1 + r(1 + \rho)]c}{\rho - \delta},
\]
then the critical expected mean return in the model of equations (1) - (2) is given as
\[
\mu^* = \frac{-(1+r)c}{1+\rho c}/[(\rho-\delta)z].
\]

Proof. If \( z \geq -(1+\rho)\mu^*/(\rho-\delta) \), then the bounds in \( B(\mu^*) \) and \( C(\mu^*) \) are interior. Hence \( B(\mu^*) = 0 \) given the constant density and the symmetry in equation (10), and \( C(\mu^*) = -2c\rho\mu^*/[(\rho-\delta)z] \) from equation (11). Equation (9) then implies equation (13). Equations (12) and (13) in turn imply the premise that \( z \geq -(1+\rho)\mu^*/(\rho-\delta) \).

Note that from equation (9) the assumption of the uniform distribution may lead to a more strongly negative value for \( \mu^* \) because it causes \( B(\mu^*) \) to be equal to its minimum value of zero. Equation (12) guarantees that the extreme realizations for the return innovations are large enough in absolute value that, in the range \( \mu_t \in [-\mu^*, \mu^*] \), a sufficiently large innovation can always occur for which a switch becomes immediately optimal for one of the two possible zero-investment positions. This assumption in the uniform distribution case replaces the assumption of unbounded support in the general case.

Equation (13) states that, given \((\rho - \delta)\) fixed, the absolute value of the optimal filter as a fraction of the transaction cost (plus interest) depends negatively on the persistence of the mean of the return process, \( \rho \): the trader should be willing to switch his position more readily toward a profitable opportunity if it is likely to persist longer. Considering equation (4) and the fact that for the uniform distribution \( z = \sqrt{3} \sigma \), the optimal filter depends positively on the variability of the mean of the return process, \((\rho - \delta)z\), scaled by the transactions cost, \( c \): if the mean is highly variable compared to the transaction cost, then a trader should require a higher immediate expected return before switching since there is a
higher chance that he may want to switch back soon.

It is instructive to evaluate a few special cases in equation (13). First, if we have a pure MA(1) process, $\rho = 0$, then the optimal filter becomes the naïve filter (plus interest), $\mu^* = -c(1 + r)$. The reason is that there is no persistence in the profit opportunities because the current conditional expected return does not depend on past conditional expected returns. Second, if we have a pure AR(1) process, $\delta = 0$, then the optimal filter becomes $\mu^* = -c(1 + r) /[1 + (c / z)]$, which is closer to zero than in the MA(1) case, implying that a switch occurs more readily (because the conditional expected return is persistent). Interestingly the filter in the AR(1) case does not depend on $\rho$. The explanation is that, as $\rho$ increases, two offsetting effects occur: due to the $\rho \mu_t$ term in equation (4) the expected return becomes more persistent, making the critical expected return less negative, but due to the $\rho \epsilon_t$ term in equation (4) the expected return also changes more rapidly, making the critical expected return more negative.

II. Empirical Illustration for Foreign Exchange Trading: Optimal Filter Calculation

The model developed in the preceding section can be interpreted in three different ways. First, if we ignore the underlying $x_t$ process in equation (2), the $\mu_t$ in equation (4) may be interpreted as an excess return that is fully known at time $t$. Thus, we are dealing with a case of pure arbitrage where the trader optimizes the after-transaction-cost return $V(\mu_t)$. Second, we may think of $\mu_t$ as the risk-adjusted expected return, the “alpha”, so that $V(\mu_t)$ represents the expected net present value adjusted for systematic risk. Similarly, $\mu_t$ could represent a particular expected utility level, which would also account for risk. Third, we can interpret $\mu_t$ as an expected return in a case where risk is relatively small or non-systematic. In this case, an appropriate risk correction can simply be applied to the ex post returns. A minor drawback is that the optimal filter has to be applied with unadjusted returns, but this is not a major issue if the risk adjustment is small and would anyway bias results away from finding positive trading
Empirically, it is difficult to find accurate data to examine the first interpretation, while the second interpretation requires employing a particular risk model. Accordingly, we adopt the third interpretation of the theory in considering uncovered interest speculation in the dollar-yen spot foreign exchange market.

A. A Parametric Moving Average Trading Strategy

As discussed extensively in the literature, see for instance Cornell and Dietrich (1978), Frankel and Froot (1990), LeBaron (1998, 1999), Lee and Mathur (1996), Levich and Thomas (1993), Qi and Wu (2006), Sweeney (1986), and others, profitable trading strategies in foreign exchange markets traditionally have employed moving-average (MA) technical trading rules. MA trading rules of size $N$ work as follows: calculate the moving average using $N$ lags of the exchange rate. Buy the currency if the current exchange rate exceeds this average; short-sell the currency if the current exchange rate falls short of this average.

Defining $s_t$ as the log of the current-period spot exchange rate level (dollar price per yen) and $\Delta s_t$ as the percentage appreciation of the yen, the implicit exchange rate forecasting model behind the MA trading rule is

$$\Delta s_{t+1} = \lambda \left( s_t - \left[ \sum_{i=1}^{N} s_{t-i} \right] / N \right) + \epsilon_{t+1}, \quad E_t \epsilon_{t+1} = 0, \quad \lambda > 0. \quad (14)$$

For any positive $\lambda$, equation (14) implies a positive expected exchange rate appreciation if the log of the current exchange rate exceeds the $N$-period MA. Empirically, we find $\lambda$ to be positive in all our specifications. Hence, the decision rule based on equation (14) to buy (short) the currency if the expected appreciation is positive (negative) leads to a trading strategy equivalent to the MA trading strategy for any

---

7 Neely (2003) finds that optimizing rules based on an ex ante risk criterion provide substantially different results compared to results adjusted for risk ex post. He notes, however, that the ex ante risk adjustment implies higher (adjusted) returns from technical trading.

8 The source of the excess returns from MA strategies in foreign exchange markets may be due to central bank intervention designed to smooth exchange rate fluctuations. See for instance Sweeney (2000) and Taylor (1982). However, recent results by Neely and Weller (2001) and Neely (2002) argue against this perspective.
positive value of $\lambda$. In what follows, our strategy is to transform equation (14) to infer empirically a value for $\lambda$ which, in addition to implying identical switch points as the MA rule, also provides a quantitative estimate of the expected gain from switching that may be compared directly to the transactions cost.

The true distribution of the exchange rate can be very complex and the simple MA process in equation (14) can only be an approximation of the true exchange rate process. We are motivated to use the MA rule because it is the most popular rule studied by researchers and used by practitioners. It is important to emphasize here that the primary purpose of this paper is not per se to search for the best exchange rate forecasting model to generate trading profitability, or to explain the potential profitability from certain trading strategies. But rather, the key point we want to make is that, given a data-generating process which exhibits some return predictability, an optimal transaction filter can be designed to maximize the after-cost expected profitability of a particular trading strategy. The optimal filter size is conditioned on a specific exchange rate forecasting model and can be easily and unambiguously computed using prior data. It is shown empirically below that the optimal filter can be significantly smaller than the naïve filter equal to the transaction cost. The optimal filter will in general outperform the naïve filters regardless of the specific return-generating process assumed.

Equation (14) can straightforwardly be rewritten as an autoregressive process in the percentage change in the exchange rate, with the coefficients in the autoregression given by the Bartlett weights:

$$\Delta s_{t+1} = \lambda \left( \sum_{i=0}^{N-1} \frac{N-i}{N} \Delta s_{t-i} \right) + \varepsilon_{t+1},$$  \hspace{1cm} (15)

Typical studies on technical analysis of foreign exchange do not utilize information on interest rates in computing the moving averages and do not estimate a parametric model for forecasting. To be fully consistent with our theory, we want to treat the excess return as the variable to be forecasted in the

---

9 Given equation (14) there is a clear link between the popular MA filters with a band and our filter. An exchange rate such that the moving average exchange rate is exceeded by x percent induces a switch. In our case, for a naïve filter equal to c, for instance, we need $\lambda x > 2c$ to induce a switch. So, for the numbers we find in our empirical section for the MA(21) process, our naïve filter corresponds to $x = 2c / \lambda = 0.1/0.025 = 4$, a 4 percent band for the ad hoc MA filter.
forecasting equation. To do so, we add the interest rate differential to the percentage change in exchange rate, so that Equation (15) becomes:

\[ x_{t+1} = \lambda \left( \sum_{i=0}^{N-1} \frac{N-i}{N} x_{t-i} \right) + \varepsilon_{t+1}, \]  (16)

where \( x_t \equiv \Delta s_t + r^j_{t-1} - r^U_{t-1} \), \( r^j_{t-1} \) is the daily Japanese interest rate, and \( r^U_{t-1} \) is the daily U.S. interest rate.

In other words, \( x_t \) denotes the excess return from buying the Japanese yen and shorting the U.S. dollar (or the deviation from uncovered interest parity).\(^{10}\)

In the presence of transaction costs, the MA rule needs to be supplemented with a filter that indicates by how far the current spot rate must exceed (or fall short of) the MA in order to motivate a trade. The advantage of equation (16) is that it is parametric and, given an estimate for \( \lambda \), can provide a quantitative measure of the filter based on comparing expected return to the transaction cost.

To obtain analytically the optimal filter for the MA criterion in the context of the model it is necessary that equation (16) be translated to ARMA(1,1) format, for which the model provides the optimal filter (which is in closed form expression if the \( \varepsilon_i \) are uniformly distributed). LeBaron (1992), Taylor (1992), and others have shown that an ARMA(1,1) well replicates moving average trading rule results. The theoretical restriction on the ARMA(1,1) process in equation (2) that \( \rho > \text{Max} (0, \delta) \) is satisfied for all our empirical specifications. Inverting the ARMA(1,1) process yields an alternative autoregression:

\[ x_{t+1} = (\rho - \delta) \sum_{i=0}^{\infty} \delta^i x_{t-i} + \varepsilon_{t+1} \]  (17)

Comparing equations (16) and (17) it follows that equation (17) is a good approximation for equation (16) if we set \( \rho - \delta = \lambda \) and if the \( (N-i)/N \) terms are close to \( \delta^i \) for all \( i \). Taking a natural log approximation, and choosing \( \delta \) to match the Bartlett weights:

\[ -i/N \approx \ln[(N-i)/N] = i \ln(\delta) \Rightarrow \delta = \exp(-1/N). \]  (18)

\(^{10}\) Deviations from uncovered interest parity and their persistence are documented in previous research. See, for example, Canova (1991), Engel (1996), and Mark and Wu (1998).
Thus, for a given $N$, we run regression (16) to obtain an estimate of $\lambda$. Then from $\delta = \exp(-1/N)$ and $\rho = \lambda + \delta$, we can uniquely identify a $\delta$ and $\rho$ that provide a good ARMA(1,1) proxy for an MA-based process with a relatively large $N$. In turn, the ARMA parameters allow us to calculate analytically the critical expected return $\mu^*$ governing the transaction choice.

B. Preliminary Dollar-Yen Process Estimates

Our data on the Japanese yen – U.S. dollar spot exchange rate cover the period from August 31, 1978 to May 3, 2003 with 6195 daily observations. Daily exchange rate data for the Japanese yen are downloaded from the Federal Reserve’s webpage. For interest rates we obtain the Financial Times’ Euro-currency interest rates from Datastream International. The first 1/3 of the sample (2065 observations) is used for model estimation. Out-of-sample forecasting starts on November 28, 1986 until the end of the sample (4130 observations). We estimate the exchange rate dynamics in four ways: 1. an ARMA(1,1) process (equation 2); 2. an AR(1) process ($\delta$ in equation 2 is set equal to zero); 3. a process consistent with an MA rule of 21 lags (21 trading days in a month), as is commonly considered with daily data (equation (16)); and 4. a process consistent with an MA rule of size 126 (half a year), around the size often used by traders, although results appear to depend little on the size of the MA process chosen (LeBaron (1998, 1999)). We choose these four models as our empirical illustration for the following reasons. Model 1, the ARMA(1,1), is the exact model assumed in our theoretical derivation, while Model 2, the AR(1) model, is a more parsimonious specification that is an interesting special case. Models 3 and 4 are commonly employed by academia and practitioners.

Columns (1)-(5) of Table I show the results of the in-sample model estimation using the first 1/3 of our sample for each way of capturing the exchange rate dynamics. For the ARMA(1,1) process we find $\rho = 0.918$ and $\delta = 0.880$ with a standard error of $\sigma_{\varepsilon} = 0.00658$; for the AR(1) process we find low persistence with $\rho = 0.0548$ and a standard error of $\sigma_{\varepsilon} = 0.00659$. Thus, both processes provide similar
accuracy although the parameters differ substantially. While the data cannot tell us clearly whether the AR(1) or the ARMA(1,1) process is better at describing the exchange rate dynamics, we will see that the implications for optimal trading are substantially different. For the representative moving average rule with 21 lags we find for the slope in equation (16) that $\lambda = 0.0257$ and $\sigma_\varepsilon = 0.00656$. Since we have $N = 21$ we obtain from equation (18) that $\rho = 0.979$ and $\delta = 0.954$, which is not statistically distinguishable from the direct estimates from the ARMA(1,1) model. The MA(126) process yields $\lambda = 0.00818$ and $\epsilon \sigma_\varepsilon = 0.00635$, implying by approximation that $\rho = 0.999$ and $\delta = 0.992$.

We assume one round trip transaction cost of $2c = 0.001$ (10 basis points) per switch throughout. Sweeney (1986) finds a transaction cost of 12.5 basis points for major foreign exchange markets, but more recent work by Bessembinder (1994), Melvin and Tan (1996), and Cheung and Wong (2000) finds bid-ask spreads for major exchange rates between 5 and 9 basis points. To account for transaction costs in addition to those imbedded in the bid-ask spread, related to broker fees and commissions and the lending-borrowing interest differential, we use 10 basis points as a realistic number for the dollar-yen market. The daily U.S. interest rate is on average over the first 1/3 of the sample equal to 0.000439 percent. This average rate is used as a proxy for the discount rate $r$ in computing the optimal filter in equation (13).

---

11 Previous more elaborate research on this issue by LeBaron (1992) and Taylor (1992) on the other hand finds that, while the ARMA(1,1) formulation does well, the AR(1) case is much poorer in replicating the key features of exchange rate series.

12 Balvers and Mitchell (1997) raise a similar issue in the context of optimal portfolio choice under return predictability.

13 Charles Engel pointed out to us the apparent puzzle that the ARMA(1,1) model $x_t - 0.918x_{t-1} = \varepsilon_t - 0.880\varepsilon_{t-1}$ implies a substantially different optimal filter from our AR(1) model, $x_t = 0.055\varepsilon_{t-1} + \varepsilon_t$. This is surprising, given that the AR(1) model can be rewritten as an ARMA(2,1) model $x_t - 0.918x_{t-1} + 0.047x_{t-2} = \varepsilon_t - 0.863\varepsilon_{t-1}$, which is very similar to the ARMA(1,1) model. The AR(1) and ARMA(1,1) models are similar in that they unconditionally describe the data with roughly the same degree of precision as indicated by $\sigma_\varepsilon$ in our Table I. The reason they imply quite different filters is because the optimal filter depends on the persistence of the conditional expected return. Equation (4) says that the conditional expected return is an AR(1) process, which becomes $\mu_t = 0.918\mu_{t-1} + (0.918 - 0.880)\varepsilon_{t-1} = 0.918\mu_{t-1} + 0.038\varepsilon_{t-1}$ for our empirical ARMA(1,1) model, while our empirical AR(1) model (or the equivalent ARMA(2,1)) implies $\mu_t = 0.055\mu_{t-1} + 0.055\varepsilon_{t-1}$. We can see that the conditional expected return from the ARMA(1,1) model is much more persistent than that from the AR(1) model. Therefore, we apply a smaller filter for the ARMA(1,1) model since even a small positive expected one-period return is enough to make up for the transaction cost because the new position is likely to remain optimal for a number of future periods.

14 Note that $c$ in the model represents the cost of closing out a zero-investment position. For most assets this requires both a sale and a purchase implying a round trip transaction cost. However, in the case of foreign exchange we directly purchase one currency with the other, implying only a one-way transaction cost. Reversing the position requires double that transaction cost $2c$ which is one round trip.
The true distribution of the exchange rate can be quite complex, and we do not know a priori which distributional assumption is the best approximation. Therefore we choose to estimate the optimal filter $\mu^*$ using three methods. First, under the assumption that the error term $\varepsilon_t$ is uniformly distributed the optimal filter, denoted $\mu^*_U$, can be analytically calculated using equation (13). Second, $\varepsilon_t$ is assumed to follow a normal distribution. In this case, the result in equation (13) no longer holds, and we estimate the optimal filter, denoted $\mu^*_N$, through Monte-Carlo simulation. Last, we do not make an assumption about the distribution of $\varepsilon_t$ and estimate the optimal filter, denoted $\mu^*_B$, by bootstrapping the model residuals $\hat{\varepsilon}_t$ with replacement.

C. The Optimal Filter Implied by the Theory under the Uniform Distribution

Under the assumption of a uniform distribution, we can obtain $z$ from the relation $z = \sqrt{3} \sigma_\varepsilon$. All the information now is there to allow us to calculate the optimal filter from equation (13) for the dollar-yen exchange rate. Column (6) of Table I provides the results. For the AR(1) case we find that the ratio of the critical return to the transaction cost is $-\mu^*_U / c = 0.92$.\textsuperscript{15} Hence, in this case the optimal filter is not very different from a naïve filter that equals the transaction cost $c$. The main reason is that, from equation (4), the persistence in the mean return is small at $\rho = 0.0548$ so that, no matter what the current holdings are, there is not much difference in future probabilities of trading.

For the 1-month MA process, the parameters backed out from the MA(21) model yield $-\mu^*_U / c = 0.23$. Note that inequality (12) is violated, as is necessary here when $-\mu^*_U / c < 0.50$, implying that the analytical value obtained from equation (13) is no longer accurate and must be viewed as a good approximation; hence it is more precise to state that $-\mu^*_U / c \approx 0.23$. Intuitively, the slow adjustment in the conditional mean for these parameter values implies that, in some cases, even the most extreme realization

\textsuperscript{15} This ratio is the immediate gain of switching to $-\mu^*$ from $\mu^*$ (the gain of $-2\mu^*$) divided by the transaction cost $2c$. 

18
of the exchange rate innovation would not be sufficient to induce switching. Hence, one would be certain of avoiding transaction costs for at least one period (and likely more) by buying/keeping the exchange with the positive expected return. This explains the low value of the critical expected return relative to the transaction cost.

The 6-month MA process, MA(126) yields the smallest filter, \( -\mu_\bullet / c = 0.083 \). One reason is the high persistence of expected return (the implied persistence parameter \( \rho = 0.999 \)). Another is the fact that by nature the long MA process is very smooth so changes in the mean occur very slowly so that the number of transactions is small, even when there is no filter. This is a possible reason for the popularity of this particular trading rule with practitioners.

For our main specification, the ARMA(1,1) case with \( \rho = 0.918 \) and \( \delta = 0.880 \), we find that the optimal filter is \( -\mu_\bullet / c = 0.32 \). The reason that this number is so much lower than under the AR(1) case is clear from equation (4). The persistence is not only high now with \( \rho = 0.918 \) but it is also high relative to the innovation in the conditional mean, given by \((\rho - \delta)\sigma_\varepsilon = 0.038 \sigma_\varepsilon \). Hence, it is highly likely that the exchange position (dollar or yen) with the currently positive expected return is going to be unchanged in the nearby future.

D. The Optimal Filter Obtained Numerically

As a check on the dependence of the results on the uniform distribution, we also find the optimal filter numerically using a Monte Carlo approach, assuming normality, and a bootstrapping approach.

For each Monte-Carlo trial, we simulate expected returns \( \mu_t \) using equation (16) with parameters estimated from the first 1/3 of the sample. We then choose the filter \( \mu_\bullet^* \) which maximizes the after-cost average excess return. This process is replicated 500 times. Column (7) of Table I reports the median value of the optimal filter to transaction cost ratio, \( -\mu_\bullet^* / c \), over the 500 Monte-Carlo trials. For each model, the ratio \( -\mu_\bullet^* / c \) is quite close to the optimal ratio implied under the uniform distribution \( -\mu_\bullet / c \).
with the difference between them never exceeding 5 percent of the transaction cost.

The actual distribution of $\epsilon_t$ may be neither uniform nor normal. In this case, we re-sample with replacement the fitted residuals $\hat{\epsilon}_t$ of equation (16) and use model parameters to generate expected return observations $\mu_t$. Similar to the Monte-Carlo experiment, for each bootstrapping trial, the optimal filter is chosen to be the one which maximizes the after-cost average excess return. Column (8) of Table I reports the median estimate of the optimal filter to cost ratio $-\mu_{\hat{\mu}}^*/c$ over 500 bootstrapping replications. Encouragingly, the optimal filters, for the theoretical uniform distribution case and the numerical normal and bootstrapping cases, are quite similar for each of the returns processes. Thus, the optimal filter value is robust to distributional assumptions.

Applying the optimal filters to the data in trading on deviations from uncovered interest parity, we expect straightforwardly that the optimal filters will outperform naïve filters. In particular, we expect that the optimal filter does better than the naïve “0” filter that is used implicitly when transaction costs are ignored for trading decisions (but not for calculating returns), because it saves on transaction costs; and does better than the naïve “c” filter that is employed when trading costs are considered myopically, because it does not filter out as many profitable transactions.

E. The Filter Obtained Empirically

An alternative, non-theoretical, procedure would establish potential trading strategy profits after transaction costs based on a prior data sample and for a grid of ratios of the (negative of the) critical expected return to the transaction cost. We then smooth the graph of the after-transaction-cost trading strategy profits over the grid and choose the ratio, $-\mu_{\hat{\mu}}^*/c$, that maximizes the ex post smoothed profits for the prior sample and apply this filter to out of sample data. This empirical procedure uses the in-sample part of the data more intensively, not just to estimate the time-series parameters but also the empirically optimal filter. However, the chosen filter depends on the smoothing procedure, and the intensive use of the in-sample data may hurt if that sample is not representative enough of the full data.
As before, we use the first 1/3rd of the sample as the prior sample. We use the first ½ of the prior sample to obtain the initial model estimates and then roll forward to the end of the prior sample, calculating the after-cost profits for a grid of values of $-\mu/c$ over the feasible range [0, 1.2] in increments of 0.02. The profit function is then smoothed using a 5th order moving average (the average of 2 leads, the current value, and 2 lags) filter and we pick the value of $-\mu/c$ maximizing the smoothed profits as the filter $-\mu^*_c/c$, to be applied out of sample.

The filters $-\mu^*_c/c$ obtained in this way are shown in figures 1-4 (we do not report the accompanying results in Table II to save space). These filters are relatively close to the theoretical filters for the ARMA(1,1) case and the two MA cases, and generate approximately the same after-cost trading profits for these cases (albeit a bit lower for the MA(21) case). The empirical filter is quite different for the AR(1) case, around 0.4 as compared to 0.9 for the theoretical cases, and generates zero after-cost profits. However, the choice of empirical filter depends crucially on the smoothing method. For instance in this AR(1) case more smoothing generates an empirical filter above 0.8.

F. Which Optimal Filter to Use in Practice?

The above results demonstrate that the distributional assumptions make very little difference in practice in determining the optimal filter. The robustness of our theoretical result in Proposition II with respect to the choice of innovation distribution is reassuring since a uniform distribution is not exactly a good proxy for many return distributions in practice. A reason for the apparent irrelevance of this assumption may be that, in this application, we implicitly have that the first three distribution moments are the same for any innovation distribution. This occurs because the mean of the distribution is set to zero in each case, the variance is calibrated to be the same in each case, and the skewness is similar in each case because of the symmetry in the returns of zero-investment positions relative to their reverse.

The effective similarity of different return distributions in this case together with the obtained

---

16 The smoothed and unsmoothed graphs of the after-cost profits for the four cases (ARMA(1,1), AR(1), MA(21), and MA(126)) are available from the authors.
similarity in the results in Table II for the different distributions both argue for using Proposition 2, which holds for the uniform distribution only, in practice. Calculation of the optimal filter can be performed directly from equation (13) and does not require simulations. Given the transaction cost and riskless interest rate, only a prior estimate of the ARMA(1,1) process is needed, yielding the autoregressive parameter, \( \rho \), the moving average parameter, \( \delta \), and the estimate of the innovation variance, \( \sigma^2 \) (the latter determining \( z \) from \( z = 3^{1/2} \sigma_e \)). Note also that the alternative models, the AR(1) and the MA models, are simply special cases of our model that we have employed to illustrate the implications of particular trading rules often used in foreign exchange markets. Based also on the results of LeBaron (1992) and Taylor (1992), who find that this specification is preferred with data prior to our holdout sample, it should be clear that the ARMA(1,1) process assumed in the theoretical model is the preferred specification for application. Accordingly, for given prior and holdout samples, the optimal filter can be obtained simply and uniquely.

III. Out-of-Sample Optimal Switching Strategy Results

We start our first-day forecast on November 28, 1986 (after the first one-third of the sample).\(^{17}\) For each of the four exchange rate return specifications, we estimate the model parameters using all observations for the first one third of the sample (up to November 27, 1986) and make the first forecast (for November 28). If the forecasted return (recall that the return is defined as the difference between the return from holding the Japanese yen, which is the percentage exchange rate change plus the one-day Japanese interest rate, and the return from holding the U.S. dollar, which is the one-day U.S. interest rate) is positive, we take a long position in the Japanese yen, and simultaneously take a short position in the dollar. Conversely, if the forecasted return is negative, we take a long position in the dollar and a short position in the yen. The difference in returns between the long and short positions represents the return from a zero-cost investment strategy. While daily data are employed in this study, bid-ask bounce is not an

---

\(^{17}\) Our results appear to be quite robust to the starting point of the forecast period: results for each of the four models are very similar if we start the forecast period at 1/4 or 1/2 of the sample instead of at 1/3.
issue here since the exchange rate data give the last observation on the *midpoint* of the bid and ask prices; further, due to the heavy trading volume of the Japanese yen, non-synchronous trading issue is not relevant. Additionally, while the daily interest observations do not coincide exactly with the exchange rate observations, the high volatility of the exchange rate relative to the interest rates implies that any bias due to a timing mismatch is probably negligible.

From the second forecasted day (November 29) until the end of the sample, our strategy works as follows. For each day, we use all available observations to estimate the model parameters and forecast the excess return for the following day. If either of the following two conditions occurs, a transaction will take place. (1) If the forecasted excess return is positive, its magnitude is larger than the transaction cost filter, and we currently have a long position in the dollar (and a short position in the yen), then we reverse our position by taking a long position in the yen and a short position in the dollar for the following day. This counts as one trade involving two one-way transaction costs. (2) If the forecasted excess return is negative, its magnitude is larger than the transaction cost filter, and the current holdings are long in the yen and short in the dollar, then we reverse our position by taking a long position in the dollar and a short position in the yen. This counts as one transaction and again involves two one-way transaction costs. If neither of the above two conditions applies, no trade takes place. The current holdings (both long and short) carry over to the following day and no transaction costs are incurred.

We compute the average excess return for the zero-cost investment strategy and the associated *t*-ratio for the out-of-sample forecasting period. We document the before-cost and after-cost excess return rates for the case *without* filter, and the after-cost excess return rates for the cases *with* transaction cost filter. For perspective, the simple buy-and-hold strategy of holding the yen and shorting the dollar over the whole out-of-sample period yields an annualized return of -0.00953 (the reverse strategy of holding the dollar and shorting the yen yields therefore +0.00953), less than one percent. This return is not statistically distinguishable from zero (*t*-ratio = 0.348).

Table II reports the results for the four forecasting models. Each model is discussed separately below. For the ARMA(1,1) model, the before-cost excess return in the case without a filter is 5.7 percent...
per annum which is statistically significant at the 5 percent level. The strategy, however, requires 444 trades over 4,130 trading days and accounting for the round-trip costs of 10 basis points reduces the after-cost excess return to 3.0 percent which is no longer statistically significant. The naïve filter equal to the 5 basis points one-way transaction cost “c” dramatically reduces the number of trades to 12, resulting in a lower excess return of 0.6 percent which is statistically insignificant. In contrast, the optimal filter, \( \mu^*_U \), captures many of the profitable trades and yields an excess return of 5.5 percent which is significant at the 5 percent level. The filter under the bootstrapped distribution \( \mu^*_B \) produces nearly the same results as \( \mu^*_U \), whereas the filter under the normality assumption, \( \mu^*_N \), yields an even higher return of 7.0 percent which is significant at the 1 percent level.

For the AR(1) model the strategy without filter involves 2,100 switches over 4,130 trading days (over 50 percent of the time). In the absence of transaction costs, the strategy produces an annualized excess return of 6.9 percent with a \( t \)-ratio of 2.486 which is statistically significant at the 5 percent level. However, the transaction cost completely wipes out the profits, resulting in a negative excess return of 5.9 percent. A naïve filter equal to the actual round-trip transaction cost of 10 basis points dramatically reduces the number of transactions to 34, and yields an insignificant excess return of 1.1 percent per annum. While it is somewhat useful, this naïve filter may be too conservative because it does not exploit the information on the persistence of expected return in the exchange rate, thereby missing a number of profitable trades. The strategy with the optimal filter \( \mu^*_U \) under the assumption of a uniform distribution captures just that opportunity. It produces 42 trades and yields a higher excess return of 5.8 percent which is significant at the 5 percent level. Similarly, the optimal filter under the normality assumption \( \mu^*_N \) produces an average excess return of 5.6 percent per annum also significant at the 5 percent level. The bootstrapped filter \( \mu^*_B \) yields an insignificant excess return of 2.6 percent.

For the MA(21) model, our strategy with the optimal filters again generates higher excess returns than the alternatives, although none of the excess returns are statistically significant at the 5 percent level.

Finally, the long MA(126) process provides a very smooth forecast of expected returns. While the
strategy without filter yields an after-cost return of 5.7 percent which is significant at the 5 percent level, the naïve filter equal to \( c \) skips too many profitable trades, resulting in a negative return of 2 percent. The optimal filter \( \mu_U^* \), while very small relative to \( c \), is capable of filtering many days with low expected returns and capturing those days when expected returns are substantial. This filter produces an expected return of 5.9 percent which is significant at the 5 percent level. The other two filters, \( \mu_N^* \) and \( \mu_B^* \) yield somewhat smaller returns (4.7 percent) which are statistically significant only at the 10 percent level.\(^{18}\)

Figures 1 through 4 display the trading strategy returns (after cost) and trading costs for the four return processes as a function of the filter value. As expected, the trading cost declines monotonically as the filter value rises. The after-cost excess return lines illustrate that in all cases the \textit{ex ante} optimal filters are reasonably close to the \textit{ex post} optimum (with the empirical filter in the AR(1) case being the only exception). Since the actual data are just one random draw from the unobserved true process this is all one should expect of a good model. Except for the AR(1) case, the trading strategy returns display the hump-shaped pattern expected for the after-cost returns.

A striking feature of these four figures is that, even though the optimal filters differ radically across the four cases, the empirical (ex post) maximum filter value is quite close to the (ex ante) optimal filter in all four cases. While each case approximates the true data process to a certain extent, it is not surprising that the ARMA(1,1) process provides the best overall fit as it is well-known to be a parsimonious description of general ARMA(p,q) processes. The strong performance of the ARMA(1,1) process and the poorer performance of the AR(1) process is consistent with the results of LeBaron (1992) and Taylor (1992) that ARMA(1,1) processes are far better at capturing the key features of exchange rate series.\(^{19, 20}\)

\(^{18}\) The trading strategies for each of the forecasting models imply a reasonably even choice of each currency. For instance, with the optimal filter \( \mu_U^* \), the fraction of long Japanese yen and short U.S. dollar is: ARMA(1,1) 1830/4130, AR(1) 2555/4130, MA(21) 1987/4130, and MA(216) 1875/4130.

\(^{19}\) Figures available from the authors provide a breakdown of the effect of the optimal filters on trading frequency. In each of the models, the optimal filters, the \( \mu_U^* \), reduce trading frequency considerably but the trades remain quite evenly distributed over time. For instance, for the ARMA(1,1) model, a minimum of two trades and a maximum of eight trades occurs in each (full) year under the optimal filter trading strategy.

\(^{20}\) A table available from the authors provides risk-adjusted trading rule returns. We correct the \textit{ex post} trading rule returns
IV. Conclusion

If transitory profitable trading opportunities exist, filter rules are used in practice to mitigate transaction costs. The filter size is difficult to determine \textit{a priori}. This paper uses a dynamic programming framework to design a filter that is optimal in the sense of maximizing expected returns after transaction costs. The optimal filter size depends negatively on the degree of persistence of the profitable trading opportunities, positively on transaction costs, and positively on the standard deviation of shocks.

We apply our theoretical results to foreign exchange trading by parameterizing the moving average strategy often employed in foreign exchange markets. The parameterization implies the same decisions as the moving average rule in the absence of transaction costs, but has the advantage of translating the buy/sell signal into the same units as the transaction costs so that the optimal filter can be calculated.

Application to daily dollar-yen trading demonstrates that the optimal filter is not solely of academic interest but may differ to an economically significant extent from a naïve filter equal to the transaction cost. This depends importantly on the time series process that we assume for the exchange rate dynamics. In particular, we find that for an AR(1) process the optimal filter is close to the naïve transaction cost filter, but for the ARMA(1,1) process the optimal filter is only around 30 percent of the naïve transaction cost filter, and for the more stable MA processes, the optimal filter is smaller still as a fraction of the transaction cost. Impressively, the \textit{ex ante} optimal filters under the assumptions of uniform, normal, and bootstrap distributions are all very close to one another and all are quite close to the \textit{ex post} after-cost return maximizing level.

We confirm that simple daily moving average foreign exchange trading generates positive returns that disappear after accounting for transaction costs. However, when the optimal filter is used, returns after transaction costs remain positive and are higher than for naïve filters. This result has implications beyond foreign exchange markets as it cautions generally against dismissing abnormal returns as due to from each of the four forecast models for CAPM market risk using the MSCI World market index and the Euro dollar interest rate as the risk free rate (results using the U.S. S&P 500 value-weighted market index are similar). In all cases the market risk sensitivities of the zero-cost investment positions are near zero. Thus, the risk-adjusted returns, the “alphas”, are very close to the unadjusted returns.
transactions costs, merely because the after-cost return is negative or insignificant. For instance, Lesmond et al. (2004) argue convincingly that momentum profits disappear when actual transaction costs are properly considered, even after accounting for the proportion of securities held over in each period. But their after-cost returns are akin to those for our suboptimal zero filter strategy. It would be interesting to see what outcome would arise if an optimal filter were used.

In our sample the trading strategy returns, gross of transaction costs, are significantly positive, but no longer significant after transaction costs are subtracted. But if we optimally eliminate trades that do not make up for their transaction cost then the after-cost profits are only slightly lower than the gross profits from unrestricted trading and are statistically significant. They are also economically significant, around 0.5 percent per month after transaction costs, which raises the issue of market efficiency. The profits are of similar magnitude as the momentum profits after transaction costs and may in fact be closely related to the momentum phenomenon. However, given the lower total variance of the trading strategy returns\textsuperscript{21}, it is even more difficult here, compared to the momentum case for equity returns, to argue that an unobserved systematic risk is responsible. So the “anomaly” may be exploitable and, in the absence of a risk explanation, could suggest market inefficiency.

Apart from the practical advantages of using the optimal filter, there is also a methodological advantage: in studies attempting to calculate abnormal returns from particular trading strategies in which transaction costs are important, there is no guideline as to what filter to use in dealing with transaction costs. Lesmond et al. (2004, p.370) note: “Although we observe that trading costs are of similar magnitude to the relative strength returns for the specific strategies we consider, there is an infinite number of momentum-oriented strategies to evaluate, so we can not reject the existence of trading profits for all strategies.” Rather than allowing the data mining problem that is likely to arise when a variety of filter sizes are applied, our approach here provides a unique filter in equation (13) that can be unambiguously obtained in advance from observable variables.

\textsuperscript{21} For example, from Jegadeesh and Titman (1993), the standard deviation of momentum excess return for the 6-month sorting and 6-month holding strategy is 5.4 percent per month. Based on our theoretical optimal filter, $\mu^*_U$, the standard deviation of after-cost return for the ARMA(1,1) specification is 0.7 percent per day, or 3.2 percent per month which is lower than that from the momentum strategy.
Appendix

A. Proof of Proposition 1

In equation (1) consider that \( V_2(x_{t+1}, \varepsilon_{t+1}) = V_1(-x_{t+1}, -\varepsilon_{t+1}) \) by virtue of the fact that, for the symmetric distribution of innovations, the expected returns of one position are equivalent to the expected returns of the reverse position if the return variables in equation (2) are the negative. Define \( \mu_t \equiv E_{t-1} x_t \). Then, given equation (2), we have \( \mu_{t+1} \equiv \rho x_t - \delta \varepsilon_t \) and thus we can without loss of generality redefine the state variables so that \( V_t(x_t, \varepsilon_t) \equiv V(\mu_{t+1}, \varepsilon_t) \). Then equation (1) becomes

\[
V(\mu_t, \varepsilon_{t-1}) = \mu_t + E_{t-1} \left[ \max \left( \frac{V(\mu_{t+1}, \varepsilon_t)}{1+r}, \frac{V(-\mu_{t+1}, -\varepsilon_t)}{1+r} - 2c \right) \right], \tag{A1}
\]

To derive equation (4), consider equation (2), \( x_t - \rho x_{t-1} = \varepsilon_t - \delta \varepsilon_{t-1} \). Taking conditional expectations on both sides yields \( \mu_t = \rho x_{t-1} - \delta \varepsilon_{t-1} \). Now lag the first equation by one period and take conditional expectations at time \( t-2 \). Simple algebra yields \( x_{t-1} = \mu_{t-1} + \varepsilon_{t-1} \). Combining the last two equations produces equation (4), which says that the conditional expected return of the ARMA(1,1) model is an AR(1) process. It is further clear, by inspection of the decision problem now summarized by equations (A1) and (4), that, once the state variable \( \mu_t \) is considered, there is no additional role for the state variable \( \varepsilon_{t-1} \).

Thus we obtain \( V(\mu_t, \varepsilon_{t-1}) \equiv V(\mu_t) \) and (A1) becomes:

\[
V(\mu_t) = \mu_t + E_{t-1} \left[ \max \left( \frac{V(\mu_{t+1})}{1+r}, \frac{V(-\mu_{t+1})}{1+r} - 2c \right) \right]. \tag{A2}
\]

By the symmetry of the density function, we also have

\[
V(-\mu_t) = -\mu_t + E_{t-1} \left[ \max \left( \frac{V(-\mu_{t+1})}{1+r}, \frac{V(\mu_{t+1})}{1+r} - 2c \right) \right]. \tag{A2'}
\]

From (A2) and (A2'), we have

\[
V(-\mu_t) - V(\mu_t) = -2\mu_t + \int_{\{\varepsilon_t: -2c < V(-\mu_{t+1}) - V(\mu_{t+1}) \leq 2c\}} \frac{V(-\mu_{t+1}) - V(\mu_{t+1})}{1+r} dF(\varepsilon_t) + 2c \left[ \int_{\{\varepsilon_t: V(-\mu_{t+1}) - V(\mu_{t+1}) > 2c\}} dF(\varepsilon_t) - \int_{\{\varepsilon_t: V(-\mu_{t+1}) - V(\mu_{t+1}) < -2c\}} dF(\varepsilon_t) \right], \tag{A3}
\]

which evaluates the Max contingent on the possible realizations of the innovation \( \varepsilon_t \). It can be shown by induction that \( V(-\mu_t) - V(\mu_t) \) is strictly decreasing in \( \mu_t \) (it is true for some terminal time \( T \); then, if it holds for some time \( t+1 \), and given that \( \rho > 0 \) in equation (4), the second term on the right-hand side of (A3) is also decreasing in \( \mu_t \) and we can establish the result for time \( t \). Then let \( T \) go to infinity). Further, a unit decrease in \( \mu_t \) increases \( V(-\mu_t) - V(\mu_t) \) by at least 2 (the direct gain from switching). Hence, since the innovation \( \varepsilon_t \) has unbounded support, then the distribution of \( \mu_{t+1} \) is unbounded and there exists exactly
one finite value $\mu^*$ at which $\frac{V(-\mu^*) - V(\mu^*)}{1 + r} = 2c$. Accordingly, the investor maintains the current zero-
investment position whenever $\mu_{t+1} > \mu^*$ and reverses the current position whenever $\mu_{t+1} \leq \mu^*$. It is then
straightforward to convert (A2) to equation (3).

B. Derivation of Equation (5)

From equation (3), $V(-\mu_i)$ is given as

$$V(-\mu_i) = -\mu_i + \int_{-\mu_i}^{\mu_i} \left( \frac{V(-\mu_i) + (\rho - \delta)\varepsilon_i}{1 + r} \right) dF(\varepsilon_i) + \int_{-\mu_i}^{\mu_i} \left( \frac{V(\mu_i) + (\rho - \delta)\varepsilon_i}{1 + r} - 2c \right) dF(\varepsilon_i) \quad (A4)$$

Using the symmetry property of the density function and a change in variable from $\varepsilon_i$ to $-\varepsilon_i$, we can write equation (A4) as follows:

$$V(-\mu_i) = -\mu_i + \int_{-\mu_i}^{\mu_i} \left( \frac{V(-\mu_i) + (\rho - \delta)\varepsilon_i}{1 + r} \right) dF(\varepsilon_i) + \int_{-\mu_i}^{\mu_i} \left( \frac{V(\mu_i) + (\rho - \delta)\varepsilon_i}{1 + r} - 2c \right) dF(\varepsilon_i) \quad (A5)$$

Then, from equations (3) and (A5), we obtain

$$V(-\mu_i) - V(\mu_i) = -2\mu_i + B(\mu_i) + C(\mu_i),$$

with

$$B(\mu_i) = \int_{-\mu_i}^{\mu_i} \left( \frac{V(-\mu_i) - V(\mu_i)}{1 + r} \right) dF(\varepsilon_i), \quad (A6)$$

$$C(\mu_i) = -2c \left[ \int_{-\mu_i}^{\mu_i} dF(\varepsilon_i) - \int_{-\mu_i}^{\mu_i} dF(\varepsilon_i) \right] = 2c \int_{-\mu_i}^{\mu_i} dF(\varepsilon_i), \quad (A7)$$

where the last equality follows from the symmetry of the density function. Thus,

$$V(-\mu_i) - V(\mu_i) = -2\mu_i + \int_{-\mu_i}^{\mu_i} \left( \frac{V(-\mu_i) - V(\mu_i)}{1 + r} \right) dF(\varepsilon_i) + 2c \int_{-\mu_i}^{\mu_i} dF(\varepsilon_i), \quad (A8)$$

which is equation (5) given the definitions in equations (6) and (7).
References


Qi, Min, and Yangru Wu, 2006, Technical trading-rule profitability, data snooping, and reality check: evidence from the foreign exchange market, *Journal of Money, Credit and Banking* 30, 2135-2158.


Table I Model Parameters and Implied Optimal Transaction Cost Filters

This table reports parameter estimates for candidate forecasting models and for each model the implied optimal filter/transaction cost ratios. The forecasting models are ARMA(1,1): \[ x_{t+1} = \rho x_t + \epsilon_{t+1} - \delta \epsilon_t; \] AR(1): \[ x_{t+1} = \rho x_t + \epsilon_{t+1}; \] and the transformed model of MA(N): \[ x_{t+1} = \lambda \sum_{i=0}^{N-1} \left( \frac{N-i}{N} \right) x_{t-i} + \epsilon_{t+1}, \] with \( N = 21 \) and 126, where \( x_t \equiv \Delta s_t + r_{t-1}^{JP} - r_{t-1}^{US} \), \( s_t \) is the log of the U.S. dollar price of one Japanese yen, \( r_{t}^{JP} \) is the daily Japanese interest rate, and \( r_{t}^{US} \) is the daily U.S. interest rate. The full sample data cover the period from August 31, 1978 to May 3, 2003 with 6,195 daily observations. The parameters are estimated with the first 1/3 of the sample (2,065 observations). The implied optimal transaction cost filters are calculated under three differential distributional assumptions of \( \epsilon_t \): uniform where the optimal filter denoted by \( \mu^*_U \) is calculated using equation (13); normal, where the optimal filter denoted by \( \mu^*_N \) is estimated through Monte-Carlo simulation with 500 replications; and bootstrap, where the optimal filter denoted by \( \mu^*_B \) is estimated through bootstrapping with replacement with 500 replications. The round-trip transaction cost \( 2c = 0.1 \) percent. Numbers inside parentheses are t-ratios.

<table>
<thead>
<tr>
<th>Model</th>
<th>( \rho )</th>
<th>( \delta )</th>
<th>( \lambda )</th>
<th>( \sigma_\epsilon )</th>
<th>( -\mu^*_U / c )</th>
<th>( -\mu^*_N / c )</th>
<th>( -\mu^*_B / c )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARMA(1,1) Model</td>
<td>0.918</td>
<td>0.880</td>
<td>0.00658</td>
<td>0.32</td>
<td>0.34</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(19.232)</td>
<td>(15.665)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR(1) Model</td>
<td>0.0548</td>
<td>0.00659</td>
<td>0.92</td>
<td>0.88</td>
<td>0.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.494)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA(21) Model</td>
<td>0.0257</td>
<td>0.00656</td>
<td>0.23</td>
<td>0.24</td>
<td>0.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.846)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA(126) Model</td>
<td>0.00818</td>
<td>0.00635</td>
<td>0.083</td>
<td>0.12</td>
<td>0.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.518)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table II  Effects of Transaction Costs on Trading Performance in Foreign Exchange

This table reports trading performance in the Japanese yen with a round-trip transaction cost $2c = 0.1$ percent. The data cover the period from August 31, 1978 to May 3, 2003 with 6,195 daily observations. The first 1/3 of the sample (2,065 observations) is used for model estimation. The parameter estimates are used to calculate the optimal transaction cost filters. Out-of-sample forecasting starts on November 28, 1986 until the end of the sample (4,130 observations). Columns 1-5 display results where no transaction cost filter is imposed. Columns 2-3 report the before-cost excess returns (from the zero-cost investment strategies) and $t$-ratios, whereas Columns 4-5 report the after-cost excess returns and $t$-ratios. Similarly, Columns 6-8 report the results (after-cost) when a naïve filter equal to the actual transaction cost $c$ is imposed. Columns 9-11, 12-14, and 15-17, show the results when the optimal filters, $\mu_U^*$, $\mu_N^*$ and $\mu_B^*$, are imposed, respectively. All returns are annualized.

<table>
<thead>
<tr>
<th>Model</th>
<th>without transaction cost filter</th>
<th>with naïve filter “c”</th>
<th>with optimal filter $\mu_U^*$</th>
<th>with optimal filter $\mu_N^*$</th>
<th>with optimal filter $\mu_B^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>before cost</td>
<td>after cost</td>
<td>before cost</td>
<td>after cost</td>
<td>before cost</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td># of trades</td>
<td>excess return</td>
<td>t-ratio</td>
<td># of trades</td>
<td>excess return</td>
</tr>
<tr>
<td>AR(1)</td>
<td>444</td>
<td>0.057</td>
<td>2.054</td>
<td>0.030</td>
<td>1.073</td>
</tr>
<tr>
<td>MA(21)</td>
<td>2100</td>
<td>0.069</td>
<td>2.486</td>
<td>-0.059</td>
<td>-2.144</td>
</tr>
<tr>
<td>MA(126)</td>
<td>448</td>
<td>0.040</td>
<td>1.453</td>
<td>0.013</td>
<td>0.463</td>
</tr>
<tr>
<td></td>
<td>131</td>
<td>0.065</td>
<td>2.338</td>
<td>0.057</td>
<td>2.047</td>
</tr>
</tbody>
</table>
Note: The solid line with round symbols displays the after-cost excess return and the one with square symbols displays the trading cost. The vertical lines represent different filters: the solid line is the theoretical optimal filter under uniform distribution $-\mu^*_U/c$; the short dashed line is the optimal filter under normal distribution $-\mu^*_N/c$; the long dashed line is the optimal filter under bootstrap distribution $-\mu^*_B/c$; and the dotted line is the ad hoc filter which maximizes the in-sample smoothed after-cost excess return using the first one third of the sample $-\mu^*_E/c$. 

Fig 1. The Effect of the Filter on After-cost Returns for the ARMA(1,1) Model
Note: The solid line with round symbols displays the after-cost excess return and the one with square symbols displays the trading cost. The vertical lines represent different filters: the solid line is the theoretical optimal filter under uniform distribution $-\mu^*_U/c$; the short dashed line is the optimal filter under normal distribution $-\mu^*_N/c$; the long dashed line is the optimal filter under bootstrap distribution $-\mu^*_B/c$; and the dotted line is the ad hoc filter which maximizes the in-sample smoothed after-cost excess return using the first one third of the sample $-\mu^*_E/c$. 

Fig 2. The Effect of the Filter on After-cost Returns for the AR(1) Model
Note: The solid line with round symbols displays the after-cost excess return and the one with square symbols displays the trading cost. The vertical lines represent different filters: the solid line is the theoretical optimal filter under uniform distribution $-\mu_U^* / c$; the short dashed line is the optimal filter under normal distribution $-\mu_N^* / c$; the long dashed line is the optimal filter under bootstrap distribution $-\mu_B^* / c$; and the dotted line is the ad hoc filter which maximizes the in-sample smoothed after-cost excess return using the first one third of the sample $-\mu_E^* / c$. 
Note: The solid line with round symbols displays the after-cost excess return and the one with square symbols displays the trading cost. The vertical lines represent different filters: the solid line is the theoretical optimal filter under uniform distribution $\mu^*_c / c$; the short dashed line is the optimal filter under normal distribution $-\mu^*_N / c$; the long dashed line is the optimal filter under bootstrap distribution $-\mu^*_B / c$; and the dotted line is the ad hoc filter which maximizes the in-sample smoothed after-cost excess return using the first one third of the sample $-\mu^*_E / c$. 