Exchange Rate Modelling and Forecasting

by

Michael Sager

A thesis submitted in fulfillment of the requirements of
the degree of Doctor of Philosophy in Economics

University of Warwick, Department of Economics

November 2004

Abstract

The objective of this thesis is to assess the current state of exchange rate modelling and forecasting. The thesis has four distinct essays, each one analysing a current interest topic in this wide and vibrant area of economic research. But a common thread runs through all four: to determine whether it is possible to use the results of this research to develop trading strategies that can add persistent value to international investment portfolios with significant exposure to the foreign exchange market. This market has a daily turnover of $1.9 trillion (BIS, 2004) and is the most liquid financial exchange in the world, by some distance. Nonetheless, we argue that the market is also inefficient, in the sense that profitable trading opportunities persist and that prices do not reflect all available public information on a continuous basis. If we are correct-and we present simulation results that suggest we are-then the opportunity to derive and test plausible trading rules for the management of international investment portfolios though rigorous academic research is enormous. Yet all too often academic exchange rate research appears to be conducted in a cocoon, with the result that conclusions are sometimes difficult to apply in a practical context by portfolio managers. These difficulties reflect the computational requirements of implementing highly intensive trading strategies, associated trading costs and size limitations, and the practical limitations on implementation raised by publication lags and general data limitations.

We aim to address these difficulties throughout this thesis. By assessing the merits of various theoretical models that collectively encompass all of the main themes on the current research agenda, we will be in a position to appreciate both the statistical and economic value of existing academic research, isolating areas of real merit for the investment community, and suggesting areas for further attention.

For Michelle and Belle

I thank Mark Taylor for his enthusiasm, guidance and teaching. Without his efforts this project would not have been possible. The opportunity to learn from the best has truly been appreciated. I also thank Michelle for her unerring support, understanding and patience.

Contents

[1] This paper was written with Michael Melvin and Mark Taylor.

1 Introduction

And you run and you run to catch up with the sun, but it's sinking
Racing around to come up behind you again.
The sun is the same in a relative way, but you're older
Shorter of breath and one day closer to death.

`Pink Floyd, Time (The Dark Side of the Moon).`

The objective of this thesis is to assess the current state of exchange rate modelling and forecasting. The thesis has four distinct essays, each one analysing a current interest topic in this wide and vibrant area of economic research. But a common thread runs through all four: to determine whether it is possible to use the results of this research to develop trading strategies that can add persistent value to international investment portfolios with significant exposure to the foreign exchange market. This market has a daily turnover of $1.9 trillion (BIS, 2004) and is the most liquid financial exchange in the world, by some distance. Nonetheless, we argue that the market is also inefficient, in the sense that profitable trading opportunities persist and that prices do not reflect all available public information on a continuous basis. If we are correct-and we present simulation results that suggest we are-then the opportunity to derive and test plausible trading rules for the management of international investment portfolios though rigorous academic research is enormous. Yet all too often academic exchange rate research appears to be conducted in a cocoon, with the result that conclusions are sometimes difficult to apply in a practical context by portfolio managers. These difficulties reflect the computational requirements of implementing highly intensive trading strategies, associated trading costs and size limitations, and the practical limitations on implementation raised by publication lags and general data limitations.

We aim to address these difficulties throughout this thesis. By assessing the merits of various theoretical models that collectively encompass all of the main themes on the current research agenda, we will be in a position to appreciate both the statistical and economic value of existing academic research, isolating areas of real merit for the investment community, and suggesting areas for further attention. The main themes begin from an analysis of longer-term fundamental determinants of exchange rate behaviour, with a particular emphasis upon Purchasing Power Parity (PPP). We then consider the exchange rate forecasting framework proposed by Clarida and Taylor (1997) that is based upon the forward rate term structure and assess whether this approach can successfully be applied to the management of foreign exchange exposures on a weekly basis. In the final two essays we assess two of the central aspects of the burgeoning market microstructure literature that focuses upon an analysis of daily and intra-day exchange rate returns: first, the contemporaneous explanatory power and accuracy of n-step ahead forecasts derived from interdealer and customer order flow data; and second, the significance of daily and intra-day exchange rate volatility impacts of macroeconomic policy announcements. To the best of our knowledge, this is the most comprehensive study of the practical implications of theoretical exchange rate research currently available.

The remainder of this introductory section provides a brief summary of each of the four essays in turn.

1.1 Explaining the Persistence of Deviations from PPP: A Non-Linear Harrod-Balassa-Samuelson Effect?

Purchasing Power Parity (PPP) in its linear form has recently been rehabilitated as the pre-eminent explanation of long-term equilibrium real exchange rate determination (Froot and Rogoff, 1994; Lothian and Taylor, 1996; Sarno and Taylor, 2002). Over shorter time horizons PPP appears less appropriate, reflecting the stylised fact that deviations of spot exchange rates away from PPP-based equilibria are typically persistent, consistent with the presence of a unit root or near-unit root process. The extent of this persistence, measured in terms of the half-life of shocks, is traditionally estimated to lie in the region of three to five years (Froot and Rogoff, 1994). This compares with a typical international portfolio investment horizon of one to two years. Similarly, the magnitude and high volatility of PPP deviations present particular difficulties for risk-averse investors concerned with the volatility and drawdown characteristics of portfolio returns as well as the sign of these returns. As a result, the naive PPP hypothesis has limited applicability in any practical financial context. Another shortcoming of PPP from an investment portfolio context is the issue of endogeneity, and therefore causality. PPP emphasises the tendency of national price levels and the nominal exchange rate to adjust to shocks to ensure either level or trend stationarity of the real exchange rate over time. But neither element of this relationship is exogenously determined, implying that nothing can be stated a priori about the directional transmission of shocks. In an investment portfolio context, the directional causality of shocks to PPP-based equilibria is a crucial issue. Implementing portfolio management decisions on the basis of a theory that embodies two-way causality or, worse, a lead-lag relationship that runs from the exchange rate to prices, is inappropriate. It is important, therefore, that the source and persistence of deviations from PPP-based equilibria are understood and addressed.

There has been much research to correct–or at least explain–the persistence of deviations from PPP. Explanations include structural market imperfections that inhibit arbitrage towards PPP and statistical biases introduced into half-life estimates of persistence by the choice of estimation methodology. Studies that attempt to correct the length of PPP deviations typically augment existing linear PPP-based equilibrium models in two ways. First, by integrating into the basic model the impact of shocks to real variables that may explain a significant proportion of the persistence of deviations of spot exchange rates from PPP.

Consistent with the work of Beveridge and Nelson (1981), persistent shocks will typically be supply related and may incorporate the well known Harrod-Balassa-Samuelson (HBS) effect that derives from intra- and inter-economy sector productivity differentials (Harrod, 1933; Balassa, 1964; Samuelson, 1964), as well as shifts in the Terms of Trade due to changes in consumer preferences towards the output of the domestic country or to commodity price shocks that impact the production base of the domestic economy. Second, to model the dynamic relationship between the real exchange rate and its fundamental determinants as a non-linear process around a linear, or log-linear, cointegrating equilibrium.

Researchers have typically chosen to focus upon an analysis of either the validity of the HBS hypothesis (for instance, Asea and Cordon, 1994; Froot and Rogoff, 1994; Sarno and Taylor, 2002) or upon establishing the existence of a well-specified non-linear dynamic PPP relationship (Michael, Nobay, and Peel, 1997; O'Connell and Wei, 1997; Taylor and Peel, 2000; Taylor, Peel and Sarno, 2001; Kilian and Taylor, 2002; Sarno, Taylor and Chowdhury, 2002; Leon and Najarian, 2003). The main contribution of this paper is to assess the validity of estimating models for mark-dollar, yen-dollar, mark-sterling and yen-sterling for the floating rate era that combine both approaches, thereby incorporating the impact of real shocks to the exchange rate within a non-linear adjustment framework. Consistent with the observed presence of heterogeneous investors in the foreign exchange market, our chosen non-linear framework assumes that the tendency of the real exchange rate to mean revert back towards its PPP level is a function of the size of disequilibria, so that only large deviations are arbitraged rapidly by the market.

Our findings are consistent with the existing literature, in that they provide mixed readings. Although productivity and Terms of Trade shocks do appear important to the determination of exchange rates during our sample period, we find only limited evidence in favour of the HBS effect; more generally, positive productivity innovations appear consistent with a depreciation of the real exchange rate, suggesting that price levels fall to allow the economy to absorb output innovations. This finding is consistent with, inter alia, IMF (2002). The magnitude of equilibrium correction parameters within estimated linear Vector Equilibrium Correction Mechanism (VECMs) suggests that augmentation of traditional PPP models with these supply side variables can greatly reduce the persistence of real exchange rate deviations from PPP. In addition, evidence of residual non-linearity in these linear VECMs encourages us to model real exchange rate dynamics, incorporating the impact of supply side variables, as an ESTAR process. In turn, these models suggest that transition between unit root and mean reversion states embedded within the ESTAR framework occurs relatively quickly-within two years at the maximum-again suggesting that our approach is capable of substantially reducing the half-lives of PPP deviations. Although estimated ESTAR models are statistically well specified and significant, we also find that encompassing tests indicate little economic benefit is gained from replacing traditional linear VECMs incorporating supply variables with non-linear models of this form. We discuss a number of reasons that may help explain this last result, including our use of quarterly data that might have masked the inherent non-linear structure in real exchange rate series (Taylor, 2000) and a relatively short data span. Overall, therefore, although we appear to have made incremental progress towards explaining the persistence of PPP deviations in this study, we have also highlighted some possible issues with the recent thrust of fundamental-based research into this problem. The search goes on.

1.2 Trading the Forward Exchange Rate Term Structure

The quality of an exchange rate forecasting model is typically judged by academic researchers on its ability to generate persistently-and significantly-smaller out-of-sample errors than a naive random walk. Using this metric, it appears that little robust progress has been achieved by the academic forecasting community during the two decades that have followed publication of the seminal Meese-Rogoff (1983a, b) papers that found in favour of a random walk over a range of fundamental-based exchange rate models (for a recent survey, see the *Journal of International Economics,* 2003). This conclusion seems equally true for forecasting models based upon the emerging microstructural literature as for models based upon more traditional economic fundamentals.

A comparison of out-of-sample forecasting errors derived from theoretical and random walk models is not a particularly useful performance metric in the context of investment portfolio management. Indeed, it represents something of a straw man, a diversion from the principal areas of concern: determining the profitability of investment decisions based upon underlying exchange rate forecasts, and the associated volatility of excess returns. Few studies address these crucial issues in a rigorous manner, while others typically assume either unrealistic-or zero-transaction costs (Rosenberg and Farka, 2001), that investors have perfect foresight (Evans and Lyons, 2002), or that investment portfolios can be turned more frequently than is realistic for most investors other than boutique Hedge Funds or commodity trading advisors (CTAs), given liquidity management issues.

Despite the lack of rigorous academic evidence of an ability to generate exchange rate forecasts that out-perform a naive random walk model, investors have demonstrated a persistent ability to add value to portfolios through currency trading (Baldridge, Meath and Myers, 2000; Hersey and Minnick, 2000). Although these findings appear mutually exclusive, the apparent contradiction is resolved in two ways. First, the quality of academic forecasting models is judged on the size of associated Mean Absolute Forecasting Errors (MAFE) or Root Mean Square Forecasting Errors (RMSFE) relative to a naïve random walk, whereas investors are interested in the profitability of forecasting models irrespective of the size of MAFEs and RMSFEs. Second, academic researchers typically focus upon the accuracy of point exchange rate forecasts, whereas few investors pay these any heed, focusing instead upon the forecast directional path of an exchange rate; persistent forecasting accuracy of this form will achieve investment out-performance relative to an underlying benchmark index as long as the move in the exchange rate is sufficiently large to outweigh associated transaction costs and interest carry.

In this paper, we marry together these two strands of research-academic and investor-using the framework proposed by Clarida and Taylor (1997). Their work-and the subsequent extension by Clarida, Sarno, Taylor and Valente (2003)-represents the first serious contradiction of Meese-Rogoff (1983a, b) to emerge from academic exchange rate research. It is predicated on the proposition that the forward rate is not an optimal predictor of the future spot exchange rate, but that important information for the future path of the spot rate is nonetheless embedded within the forward rate term structure. Exploiting this information within a linear VECM estimated by Full Information Maximum Likelihood (FIML), they achieve a statistically significant reduction in forecast errors of the order of 50%-70% relative to a random walk for mark–dollar, yen-dollar, sterling-dollar and French franc-dollar and over forecast horizons that range from 4 to 52 weeks.

Generating forecast errors significantly smaller than a random walk model is certainly an important achievement, but does not guarantee a profitable exchange rate investment strategy. To this end, we replicate the analysis of Clarida and Taylor (1997)-confirming their results in so doing-and then develop a set of trading rules based upon the resulting forecasts that are assessed in terms of their ability to generate returns persistently in excess of a strategic benchmark return. Returns from each of the exchange rate models are examined individually, but also within an equally-weighted portfolio for evidence of diversification benefits that may result from combining the models in this simple manner. We then consider the merits of stop-loss limits that are designed to truncate the extent of negative returns from any trading strategy. We also consider various portfolio construction techniques regularly applied throughout the investment industry to assess the diversification benefits that derive from combining models into portfolios based upon efficient weights that take account of historical return and risk correlations, as well as drawdown parameters that are central to many risk averse investors in the foreign exchange market. We contrast the results of these rules with a naive Forward Rate Bias (FRB) strategy that is widely utilised throughout the foreign exchange investor community.

Using two simple trading rules, with associated portfolio construction tools, we demonstrate that the Clarida-Taylor framework can profitably be applied to investment portfolios, generating returns persistently in excess of an underlying strategic benchmark for euro-dollar, yen-dollar and sterling-dollar.[2] To the best of our knowledge these results are the first demonstration of an ability to marry together academic and investor strands of exchange rate research under realistic transaction costs, position limits and publication lags within a profitable trading strategy. Furthermore, the simulated trading results that we report for the Clarida-Taylor framework appear superior to the performance of a traditional FRB strategy that is widely applied across the investor community, and also seem more consistent with investor sensitivities to the risk and drawdown

[2]Although included in the Clarida-Taylor analysis, we do not analyse the French franc-dollar exchange rate in this study, given its replacement in 1999 by the euro-dollar rate.

characteristics of excess returns than is traditionally the case for FRB strategies.

1.3 The Role of Order Flow in Exchange Rate Forecasting

While empirical evidence over the post-Bretton Woods period suggests that fairly standard macroeconomic fundamentals–such as relative monetary velocity–may influence the long-run behaviour of real and nominal exchange rates (for surveys see Frankel and Rose, 1995; Froot and Rogoff, 1995; Taylor, 1995; Sarno and Taylor, 2002), modeling–and especially forecasting–the exchange rate over shorter horizons remains an occupational hazard of the international financial economist, since standard economic fundamentals appear to be poorly correlated with higher frequency exchange rate movements. Largely motivated by this stylized fact, a growing literature on market microstructure has emerged in recent years to suggest that the quality of fundamental-based exchange rate forecasts can be improved by resort to measures of foreign exchange order flow (defined as signed transaction volume: Froot and Ramadorai, 2001; Lyons, 2002; Evans and Lyons, 2002), as well as variables such as surveys of market sentiment or positioning (Merrill Lynch, 2003).[3].

Order flow is initiated for a variety of reasons that differ across the various participants in the foreign exchange market. These participants include corporations, central banks, asset management firms, CTAs, hedge funds, private individuals and investment bank dealers. Participants exhibit significant heterogeneity, in terms of opportunity sets and risk-return expectations, and display distinct informational asymmetries, with some participants better informed than others. By reputation, customer order flow is the primary source of private information in the foreign exchange market (Lyons, 1995; Ito, Lyons and Melvin, 1998; Bjønnes and Rime, 2001a; Rime, 2001). This private information is typically assumed to relate to future innovations in fundamental exchange rate determinants, including monetary policy innovations (Evans and Lyons, 2002; Lyons, 2003; Jansen and de Haan, 2003). But it can also incorporate knowledge of the decision-making process that triggers strategic shifts in portfolio benchmark hedge ratios in response to changes in risk appetite or return objectives independently from innovations in published fundamentals (Lyons, 2002). Similar to innovations in fundamental variables, changes in long-term investment objectives will lead to asset allocation shifts within investment portfolios, for instance between international bonds and equities, that in turn inspire order flow.

[3] Foreign exchange order flow should not be confused with transaction volume; the latter is a measure of trading activity between customers and dealers, or within the interdealer market, over a given period and in a particular exchange rate without indication of the direction of these transactions. By contrast, order flow is defined as signed transaction volume (Lyons, 2001), with the sign of a transaction determined by the initiating agent. Order flow therefore provides an indication of the relative strength of buy (sell) orders between, say, customers and dealers, with a purchase (sale) by the customer recorded as a net buy (sell). In this way, order flow within particular investor groups will not necessarily sum to zero, but can instead exhibit persistent trends if, say, customers build a long (short) position in a particular exchange rate relative to an underlying neutral benchmark position.

A central hypothesis of the microstructure literature is that order flow allows the wider market to learn about the private information and trading strategies of better informed participants, and therefore represents the conduit through which informational asymmetries become embedded within market prices (Lyons, 1995; Bjønnes and Rime, 2001b). Overall, this hypothesis seems an intuitive explanation of the process of price discovery in the foreign exchange market. If valid, it implies that customer order flow will consistently be more important to the determination of exchange rate returns than interdealer order flow. In addition, order flow generally should have greater explanatory and predictive power for exchange rate returns than fundamental variables.

The microstructure literature draws support and scepticism in equal measure. Few disagree with the central hypothesis that order flow is the mechanism by which private information is embedded in exchange rates. Much more disharmony surrounds the assessment of the practical value of this hypothesis. This paper seeks to make two main contributions. First, as a foundation to our empirical analysis we provide an extensive description of the structure of the foreign exchange market. This description focuses upon both the interaction of the main market participants and the current market infrastructure, and in our opinion represents the most comprehensive and accurate description of the foreign exchange market available. Second, with this foundation in place, and using aggregated and disaggregated customer order flow data from two major investment banks as well as the data on interdealer order flow employed by Evans and Lyons (2002), we critically evaluate the practical value of order flow data. We undertake this evaluation using regression analysis and Granger-causality tests that, first, incorporate contemporaneous information consistent with the existing literature (Evans and Lyons, 2002) and, second, respect real-time publication lags. We then compare the accuracy of in-sample, out-of-sample and long horizon forecasts constructed using order flow data with a naive random walk forecast, and calculate the significance of any differences in forecast errors using the Diebold-Mariano test statistic (Diebold and Mariano, 1995) for equality of forecast accuracy. To our knowledge this is the first study that has assessed the practical value of foreign exchange order flow data under realistic trading conditions and using data available to market participants on a real-time basis.

We conclude that the ability of data available to the wider market on a real-time basis to improve upon the forecasting accuracy of fundamental-based models is generally weak. In addition, and in contradiction with theoretical priors, we find widespread evidence of a strict Granger-causal relationship that runs from exchange rate returns to customer order flow. This result is consistent with evidence presented by Payne and Vitale (2002), Daníelsson, Payne and Luo (2002) and Froot and Ramadorai (2001). We discuss a number of factors that may explain our results. These include market share issues of sampled databases, pre-filtering and indexation of data, but also the validity of hypotheses that lie at the core of the microstructure literature. No single explanation can provide a complete answer. But as our study employs customer order flow from two major investment banks as well as interdealer order flow, for a range of exchange rates and sample periods, it seems reasonable to conclude that our results are relatively robust.

Our results do not invalidate the hypothesis that private information and persistent profit opportunities coexist in the foreign exchange market. Indeed, performance data from the currency overlay industry indicate that they do (Baldridge, Meath and Myers, 2000; Hersey and Minnick, 2000). Our results also do not invalidate other aspects of the microstructure literature, and particularly intra-day volatility studies that have achieved a demonstrable ability to predict and practically exploit significant volatility shifts associated with macroeconomic policy announcements (for instance, see below). But the results presented in this paper do suggest that outside of a few, particularly well informed investors who observe order flow data on an unfiltered, tick-by-tick basis, knowledge of customer or interdealer order flow cannot help improve the quality of exchange rate forecasting or the profitability of investment portfolio decision-making. From this perspective, we have confirmed the results of Meese and Rogoff (1983a, b): exchange rate forecasting remains a hazardous occupation even when the forecaster is equipped with order flow data.

1.4 Can an Old Lady Keep a Secret? A Microstructural Study of Policy Announcements at the Bank of England

The Bank of England (BoE), fondly known as the "Old Lady of Threadneedle Street," was granted operational independence to set its key repurchase, or 'repo', rate by the incoming Labour government in 1997 with the goal of creating policy consistent with price stability and economic growth. In practice, interest rate decisions are made by the Bank's Monetary Policy Committee (MPC), which meets for two days each monthly and issues a statement regarding interest rate decisions at noon on of the second each meeting day. This framework allows a natural laboratory setting for examining the impact of monetary policy decisions around a known time and date. Since the market knows that the interest rate announcement arrives at noon, there may be positioning prior to the announcement and news effects after the announcement that result in systematic patterns in exchange rate behaviour on MPC meeting days that differ from other days. We examine the evidence in the foreign exchange market to analyse the pattern of exchange rate changes and volatility surrounding the noon announcement.

One hypothesis to be explored is that positioning prior to the policy announcement could involve informed traders having superior information regarding the policy outcome. This need not involve information leaks of inside "secrets" from the MPC, but instead could reflect the activity of market participants adept at reading the public signals regarding the state of the economy and their interpretation of the likely MPC response to these signals. In addition, since activities directly related to each MPC meeting are spread over three different days, the analysis will include an examination of the pre-meeting briefing day, the first day of the meeting, and the second day of the meeting when the policy decision is made and publicised.

Our focus is on the response to meeting activities in the foreign exchange market, specifically the sterling-dollar exchange rate. Both daily and high-frequency, intraday data are employed in the analysis. The daily data provide a bird's eye view of the market around MPC meetings and then, given the findings from this low-frequency analysis, a microscope is taken to the data to examine exchange rate dynamics on days related to meetings. The intraday econometric framework is provided by a Markov switching model where exchange rate returns switch between a high-volatility, informed-trading state, and a low-volatility, uninformed or liquidity trading state on MPC days. A key difference from the usual Markov switching model employed in financial analysis is our incorporation of endogenous shifts in the transition probabilities where the shifts are modeled as a function of variables related to the MPC meeting and policy outcomes.

Daily data on the sterling-dollar exchange rate are employed to analyze any differences that may exist regarding exchange rate returns on the three kinds of days associated with MPC meetings. First, we examine tests of the equality of means and variances for the three different types of days versus all other days. From this analysis, we find no evidence of any statistically significant difference in means or variances across days, although the test for the equality of the variance on second meeting days versus all other days had the lowest p-value. Second, we estimate models of daily exchange rate returns to infer if information on MPC meeting days contains any explanatory power. Estimation results suggest that daily exchange rate returns are well characterized by mean zero changes and that meeting day information has no explanatory power for returns. However, evidence of strong GARCH effects in the daily returns was found, and incorporation of MPC meeting day information in the conditional variance equation revealed evidence of greater conditional volatility on second meeting days when interest rates are changed. Third, we estimate the probability of observing an extreme exchange rate event, defined as a return in excess of 2.5 standard deviations of the mean return, as a function of information related to MPC meetings. Our results suggest that the probability of observing an extreme exchange rate change increases by about 40 percent on the second day of MPC meetings when an interest rate change is announced.

The evidence from daily data suggests that only the second days of MPC meetings are different from other days in terms of exchange rate dynamics. Consequently, we examine second meeting day data in more detail using intraday exchange rate returns. A high-frequency sample of 5-minute observations over the period 7:00-17:00 London time is analysed using a Markov-switching framework. We assume that there exist two states: state 1, the high-volatility state associated with informed trading, and state 2, the low-volatility state associated with liquidity trading. We diverge from the usual non-linear regime-switching framework to model endogenous transition probabilities as a function of information regarding the meeting days. The transition probabilities are found to systematically switch on meeting days. The probability of remaining in the high volatility state is estimated to increase from 0.76 before noon to 0.93 between

12:00-13:30 on MPC days when interest rates are changed. In addition, the probability of remaining in the low volatility state is estimated to fall from 0.95 before 11:30, to 0.91 between 11:30-11:55, and to 0.29 at noon on MPC meeting days. So the evidence indicates that there is a statistically and economically significant news effect related to the noon announcement. We also test whether evidence exists of positioning during the meetings prior to the policy announcement at noon using various time dummy variables. We conclude in favour of some, limited, evidence of regime switching in terms of exchange rate volatility in the morning prior to the end of the MPC meetings. However, the implied change in probabilities does not appear economically significant. The news impact of MPC announcements appears to be much larger than any anticipation effect. So to answer the question posed in the title: Can an old lady keep a secret? The answer appears to be yes. The second day of MPC meetings is best characterized as having a strong exchange rate reaction to the news announcement at noon with little evidence of positioning during the morning period of the meeting.

References

[1] Asea, P.K. & W.M. Corden (1994), The Balassa-Samuelson Model: An Overview. Review of International Economics, vol. 2 , 191-200.

[2] Balassa, B. (1964), The Purchasing Power Parity Doctrine: A Reappraisal." Journal of Political Economy 72, 584-96.

[3] Baldridge, J., B. Meath, and H. Myers (2000), Capturing Alpha Through Active Currency Overlay, Frank Russell Company.

[4] Bank for International Settlements (2004), Triennial Central Bank Survey of Foreign Exchange and Derivatives Market Activity in April 2004. Preliminary Results. www.bis.org.

[5] Bean, C. (1999), The Formulation of Monetary Policy at the Bank of England. Bank of England Quarterly Bulletin, May, 434-441.

[6] Beveridge, S. and C.R. Nelson (1991), A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the "Business Cycle". Journal of Monetary Economics, 7, pp.151-74.

[7] Bjønnes, G.H. and D. Rime (2001a), US Exchange Rates and Currency Flows. Unpublished Mimeo Stockholm Institute for Financial Research. www.sifr.org.

[8] Bjønnes, G.H. and D. Rime (2001b), Customer Trading and Information in Foreign Exchange Markets. Working paper # 4. Stockholm Institute for Financial Research.

[9] Clarida, R.H. and M.P. Taylor (1997), The Term Structure of Forward Exchange Rate Premiums and the Forecastability of Spot Exchange Rates: Correcting the Errors. Review of Economics and Statistics, 79, 353-370.

[10] Clarida, R.H., L. Sarno, M.P. Taylor and G. Valente (2003), The Out-of-Sample Success of Term Structure Models as Exchange Rate Predictors: A Step Beyond. Journal of International Economics, 60, 61-83.

[11] Danielsson, J. and R. Payne (2002), Real Trading Patterns and Prices in Spot Foreign Exchange Markets. Journal of International Money and Finance 21 (2), 203-222.

[12] Daníelsson, J., R. Payne and J. Luo (2002), Exchange Rate Determination and Inter-Market Order Flow Effects. LSE Financial Markets Group Mimeo.

[13] Diebold, F., Lee, J-H. and G. Weinbach (1994), Regime Switching with Time-Varying Transition Probabilities. In: Hargreaves, C. (ed.), Nonstationary Time Series Analysis and Cointegration. Oxford University Press, Oxford and New York, 283-302.

[14] Diebold, F.X. and R. Mariano (1995), Comparing Predictive Accuracy. Journal of Business and Economic Statistics, 13, 253-62.

[15] Engel, C. and J.D. Hamilton (1990), Long Swings in the Dollar: Are They in the Data and Do Markets Know it? American Economic Review 80 (4), 689-713.

[16] Evans, M. and R.K. Lyons (2002), Order Flow and Exchange Rate Dynamics. Journal of Political Economy, February, 170-180.

[17] Frankel, J.A. and A.K. Rose (1995), Empirical Research on Nominal Exchange Rates, in G.M. Grossman and K. Rogoff (eds.) Handbook of International Economics, III, pp.1689-729. Amsterdam: North-Holland.

[18] Froot, K.A. and K. Rogoff (1994), Perspectives on PPP and Long-Run Real Exchange Rates. National Bureau of Economic Research Working Paper Working Paper #4952.

[19] Froot, K.A. and K. Rogoff (1995), Perspectives on PPP and Long-Run Real Exchange Rates, in G.M. Grossman and K. Rogoff (eds.) Handbook of International Economics, III, pp.1647-88. Amsterdam: North-Holland.

[20] Froot, K.A. and T. Ramadorai (2001), Currency Returns, Institutional Investor Flows and Exchange-Rate Fundamentals. State Street Associates Working Paper # 5.

[21] Hamilton, J.D. (1990), Analysis of Time Series Subject to Changes in Regime. Journal of Econometrics 45, 39-70.

[22] Hamilton, J.D. (1994), Time Series Analysis. Princeton University Press, Princeton, NJ.

[23] Harrod, R. (1933), International Economics. James Nisbet and Cambridge University Press.

[24] Hersey, B. and J. Minnick (2000), Currency Overlay. Global Pensions. February.

[25] International Monetary Fund (2002), The Rise of the Sterling Exchange Rate. In United Kingdom: Selected Issues. IMF Country Report # 02/46. March.

[26] Ito, T., R.K. Lyons and M.T. Melvin (1998), Is There Private Information in the FX Market? The Tokyo Experiment. The Journal of Finance, June, 1111-1130.

[27] Jansen, D.J. and J. de Haan (2003), Statements of ECB Officials and their Effect on the Level and Volatility of the Euro-Dollar Exchange Rate. CESifo Working Paper # 927.

[28] Journal of International Economics (2003), Empirical Exchange Rate Models. Vol. 60, 1, 1-234.

[29] Kilian, L. and M. P. Taylor (2002), Why is it so Difficult to Beat the Random Walk Forecast of Exchange Rates? Journal of International Economics, 60, 85-107.

[30] Leon, H. & S. Najarian (2003), Asymmetic Adjustment and Nonlinear Dynamics in Real Exchange Rates. IMF Working Paper WP/03/159. www.imf.org.

[31] Lothian and M. P. Taylor (1996), Real Exchange Rate Behaviour: The Recent Float From the Behaviour of the Past Two Centuries. Journal of Political Economy, vol. 104, 3, pp. 488-509.

[32] Lyons, R. K. (1995), Tests of Microstructural Hypotheses in the Foreign Exchange Market. Journal of Financial Economics, 39, 321-351.

[33] Lyons, R.K. (2002), Foreign Exchange: Macro Puzzles, Micro Tools. Federal Reserve Bank of San Francisco Annual Review.

[34] Lyons, R.K. (2003), Explaining and Forecasting Exchange Rates with Order Flows. 3^{RD} Bank of France Economic Forum. http://www.banque-france.fr/gb/fondatio/telechar/R_k_lyons.pdf.

[35] Meese, R.A. and K. Rogoff (1983a), Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample? Journal of International Economics, 14, 3-24.

[36] Meese, R.A. and K. Rogoff (1983b), The Out-of-Sample Failure of Empirical Exchange Rate Models: Sampling Error or Misspecification? in Frenkel, J.A. (ed.), Exchange Rate and International Economics, University of Chicago Press.

[37] Merrill Lynch (2003), Global FX Paper #3, July 22.

[38] Michael, P., R. Nobay, and D. A. Peel (1997), Transactions Costs and Nonlinear Adjustment in Real Exchange Rates: An Empirical Investigation. The Journal of Political Economy, Vol. 105, 4, 862-879.

[39] O'Connell, P.G.J. and S.-J. Wei (1997), The Bigger They Are , the Harder They Fall: How Price Differences Across US Cities Are Arbitraged, NBER Working Paper # 6089.

[40] Payne, R. and P. Vitale (2002), A Transaction Level Study of Central Bank Intervention on Exchange Rates. Unpublished LSE Financial Markets Group Mimeo.

[41] Rime, D. (2001), Private or Public Information in Foreign Exchange Markets? An Empirical Analysis. www.sifr.org.

[42] Rosenberg, M. and E. Farka (2001), Forward Rate Bias Strategy. Deutsche Bank.

[43] Samuelson, P. (1964), Theoretical Notes on Trade Problems. Review of Economics and Statistics 46, 145-54.

[44] Sarno, L. and M.P. Taylor (2002), The Economics of Exchange Rates. Cambridge: Cambridge University Press.

[45] Sarno, L., M. P. Taylor and I. Chowdhury (2002), Non-Linear Dynamics in Deviations from the Law of One Price: A Broad-Based Empirical Study. Centre for Economic Policy Research Discussion Paper # 3377.

[46] Taylor, A.M. (2000), Potential Pitfalls for the Purchasing Power Parity Puzzle? Sampling and Specification Biases in Mean Reversion Tests of the Law of One Price. National Bureau of Economic Research Working Paper # 7577.

[47] Taylor, M.P. (1995),The Economics of Exchange Rates. Journal of Economic Literature 13, 33-47.

[48] Taylor, M.P. and D.A. Peel (2000), Non-Linear Adjustment, Long-Run Equilibrium and Exchange Rate Fundamentals. Journal of International Money and Finance, 19, 33-53

[49] Taylor, M.P., D. Peel and L. Sarno (2001), Non-Linear Mean Reversion in Real Exchange Rates: Towards a Solution to the Purchasing Power Parity Puzzles. International Economic Review, 42, 1015-42.

2 Explaining the Persistence of Deviations from PPP: A Non-Linear Harrod-Balassa-Samuelson Effect?

2.1 Introduction

Purchasing Power Parity (PPP) in its linear form has recently been rehabilitated as the pre-eminent explanation of long-term equilibrium real exchange rate determination (Froot and Rogoff, 1994; Lothian and Taylor, 1996; Sarno and Taylor, 2002). Over shorter time horizons PPP appears less appropriate, reflecting the stylised fact that deviations of spot exchange rates away from PPP-based equilibria are typically persistent, consistent with the presence of a unit root or near-unit root process. The extent of this persistence, measured in terms of the half-life of shocks, is traditionally estimated to lie in the region of three to five years (Froot and Rogoff, 1994).

Persistent divergence from equilibrium indicates that linear PPP-based fundamental exchange rate models have low power to generate accurate point forecasts of the short-term dynamic path of spot around equilibrium (Meese and Rogoff, 1983a, b). These findings have encouraged much research to correct–or at least explain–the persistence of deviations from PPP. Most recently, research has focused upon an analysis of foreign exchange market microstructure, and particularly the information content of order flow data for high frequency exchange rate returns (Evans and Lyons, 2002; Danielsson, Payne and Luo, 2002; Sager and Taylor, 2004), and the behaviour of technical, or chartist, investors within the foreign exchange market (Sarno and Taylor, 2001).

As discussed in an earlier chapter, a central hypothesis of the market microstructure literature is that order flow is the mechanism by which dispersed private information is embedded in prices within the foreign exchange market. A number of researchers have reported a significant contemporaneous correlation between interdealer order flow and exchange rate returns (Evans and Lyons, 2002; Danielsson, Payne and Luo, 2002). But the absence of comprehensive, unfiltered and timely publicly available data makes this approach difficult to exploit on a real-time basis within investment portfolios.

The activity of trend-following technical investors in the foreign exchange market is also an attractive and plausible explanation for the presence of a unit root or near-unit root process in real exchange rates, particularly for observations close to Fair Value. But much of the toolkit of this investor group, including Elliot Wave and support-resistance analysis, does not lend itself to strict quantification and therefore rigorous empirical appraisal. For a comprehensive survey of technical analysis, see Sarno and Taylor (2001).

A more tractable, but not mutually exclusive, approach is to augment existing linear PPP-based equilibrium models in two ways. First, to integrate into the basic model the impact of shocks to real variables that can explain at least part of the persistence of deviations of spot exchange rates from PPP-based equilibria. Consistent with the work of Beveridge and Nelson (1981), persistent shocks will typically be supply related and may incorporate the well known Harrod-Balassa-Samuelson (HBS) effect that derives from intra- and inter-economy sector productivity differentials (Harrod, 1933; Balassa, 1964; Samuelson, 1964), as well as shifts in the Terms of Trade. Second, to model the dynamic relationship between the real exchange rate and its fundamental determinants as a non-linear process around a linear, or log-linear, cointegrating equilibrium.

Researchers have typically chosen to focus upon an analysis of either the validity of the HBS hypothesis (for instance, Asea and Cordon, 1994; Froot and Rogoff, 1994; Sarno and Taylor, 2002) or upon establishing the existence of a well-specified non-linear dynamic PPP relationship (Michael, Nobay, and Peel, 1997; O'Connell and Wei, 1997; Taylor and Peel, 2000; Taylor, Peel and Sarno, 2001; Kilian and Taylor, 2002; Sarno, Taylor and Chowdhury, 2002; Leon and Najarian, 2003). Few researchers have sought to combine both strands within one exchange rate model; see Lothian and Taylor (2004) for an exception to this general rule. Consequently, the main contribution of this paper is to assess the validity of estimating models for mark-dollar, yen-dollar, mark-sterling and yen-sterling that incorporate the impact of real shocks to the exchange rate within a non-linear adjustment framework. Inclusion of two non-dollar cross rates is intended to help avoid cross-sectional dependence that adversely affects empirical studies that focus on a set of exchange rates with a common numéraire country (O'Connell, 1998). Although this criticism is most relevant to studies that exploit panel data, where the power of traditional test statistics can be weakened substantially by the presence of this bias, inclusion of a non-dollar exchange rate will nonetheless help demonstrate the rigour of our results.

The remainder of the paper is organised as follows. In the next section we examine the various alternative approaches to fundamental exchange rate modelling with a concentration upon a discussion and explanation of the shortfalls of the PPP hypothesis.[4] In the following section we assess the validity of augmenting linear PPP models with variables that capture shocks to the real economy. A non-linear augmented PPP framework is then explored and discussed. The final section draws conclusions and suggests issues for future research.

2.2 Determinants of Equilibrium Real Exchange Rates

The estimation and interpretation of equilibrium exchange rates is a vibrant and contentious area of research. Beginning with Cassel (1918), a vast literature has developed on the merits of PPP as an appropriate explanation of the determination of equilibrium exchange rates. Following a revival of interest in PPP during the last several decades, this theory is typically now considered the benchmark against which other equilibrium theories are compared.

[4] Unless otherwise stated, we use the term PPP to describe absolute PPP. We differentiate between absolute and relative PPP below.

2.2.1 Purchasing Power Parity

In its strict form, PPP states that variation in the ratio of national price levels, expressed relative to an arbitrary base year, will exactly offset changes in the nominal exchange rate so that the real exchange rate remains constant through time. Assuming that baskets of goods used in national price indices are identical between countries, and that these goods are combined into indices using the same weights and methodology, gives

$$S = P/P^*, and\ Q = SP^*/P \tag{1}$$

where S is the domestic price of foreign currency, P and P^* are the domestic and foreign national price levels, as measured by either consumer or wholesale price indices, and Q is the real exchange rate. In testable form and logarithms, this equates to

$$s_t = \alpha + \beta(p_t - p_t^*) + \varepsilon_t \tag{2}$$

with the null hypothesis $\beta = 1$. Less strict interpretation of the PPP hypothesis allows the real exchange rate to deviate from a constant value, albeit in a (trend) stationary manner. This alternative can be written as

$$s_t = \alpha + \beta_1 p_t - \beta_2 p_t^* + \varepsilon_t \tag{3}$$

where the parameters $\beta_1$and β_2 are allowed to deviate systematically from unity over time, but s, p and p^* are assumed to exhibit a common stochastic trend, and therefore cointegrate; equivalently, ε_t, an error term assumed independently and identically distributed with mean zero and variance σ, is termed a stationary, or $I(0)$, series.

There has been much empirical analysis of the validity of PPP as an equilibrium exchange rate theory, under both fixed and floating exchange rate regimes. Comprehensive surveys of this literature can be found in Froot and Rogoff (1994), Sarno and Taylor (2002) and Coakley, Flood and Taylor (2002). As these surveys suggest, views on the validity of PPP remain mixed, depending upon exchange rates analysed, monetary regimes in sway during chosen sample periods, the length of sample period, the particular econometric technique applied, and so on. But as a general conclusion, the weight of evidence suggests that over sufficiently long data spans and using appropriately powerful tests, some variant of the PPP hypothesis is a valid characterisation of real exchange rate behaviour in the very long term (Abauf and Jorion, 1990; Froot and Rogoff, 1994; Lothian and Taylor, 1996; Rogoff and Kim, 2001; Taylor, 2002; and Sarno and Taylor, 2002). Indeed, whereas most evidence against the validity of PPP has concentrated upon data from the floating rate era-thereby bringing into question the power of associated test procedures-long span studies typically report more favourable findings. For instance, Lothian and Taylor (1996) conclude in favour of PPP using annual data for the sterling-dollar real exchange rate over the sample period 1791 to 1990. Similarly, Rogoff, Froot and Kim (2001) using annual data that span 700 years for a range of commodities traded by England and the Netherlands present results favourable to PPP; indeed, for silver-the commodity for which they have they most comprehensive price data-deviations from PPP appear to be extremely small over the course of several hundred years. And Taylor (2000) using annual data over a maximum sample of 1856 to 1996 for twenty countries also finds in favour of PPP. Overall, therefore, the results of these studies indicate that over very long periods of time the real exchange rate is a stationary series and exhibits a tendency to mean-revert in the wake of an unanticipated shock.

Although deviations from PPP may not exhibit a unit root over very long spans of data, they are relatively persistent (Rogoff, Froot and Kim, 2001).[5] Furthermore, this persistence extends beyond that implied by nominal rigidities such as non-trivial arbitrage or transportation costs. Froot and Rogoff (1994) estimate that half of a shock to PPP–the so-called half-life-is reversed after approximately three to five years (see also Abauf and Jorion, 1990; Lothian and Taylor, 1996; Rogoff, 1996; Rogoff, Froot and Kim, 2001).[6] In addition, Rogoff, Froot and Kim (2001) conclude that the persistence of PPP deviations has changed little with the advent in the last thirty years of floating exchange rates, and, more recently, a marked reduction in restrictions to cross-border trade flows. But the source of deviations does seem to have shifted, from nominal prices to nominal exchange rates, and their size appears to have increased as well (Taylor, 2002). Most research into the persistence of PPP deviations is generally consistent with the findings of Froot and Rogoff (1994). Particularly interesting are studies that estimate half-lives using disaggregated price data. For instance, Crucini and Shintani (2001), on the basis of panel estimation using annual price data for 371 traded and non-traded goods in 122 cities worldwide over the sample 1990 to 2000 report half-live estimates only slightly above the range of Froot and Rogoff if prices are assumed to exhibit a common mean across all cities; however, this estimate is reduced to around one year if different mean levels between cities and non-linear adjustment is accommodated. In addition, recent work by Cashin and McDermott (2003) that corrects for the downward bias in OLS estimates by using serial correlation and heteroskedastic robust median-unbiased estimators (MUEs) for quarterly real effective exchange rate data of twenty industrial countries over the sample 1973 to 2002 is also generally supportive of Froot and Rogoff's conclusions.[7] In addition, the use of MUEs

[5] A related issue is the high volatility of exchange rate returns relative to national price levels. We do not explore this issue in this paper.

[6] Assuming a simple AR(1) formulation, the half-life of a shock may be measured as $log(0.5)/log(\beta)$ in equation (2) above.

[7] This extent of this downward bias increases as the value of the estimator approaches unity, particularly in small data samples. It reflects the use of OLS in the presence of a unit root or near unit root process that introduces a leftwards skew to the distribution of the estimator and means that the average estimate of β in equation (2) will lie below its median value. The resolution to this bias using a MUE is first to estimate a traditional PPP regression, recovering the mean estimate of β, termed $\hat{\beta}$. MUE does not use this value of the least-squares estimator, but instead selects the value of $\hat{\beta}$ that results in a median value equivalent to the initial value of β. The probability of this estimator exhibiting either upward or downward bias is consequently equivalent, and equal in both cases to 0.50.

allows Cashin and McDermott to estimate 90% confidence intervals associated with point estimates of true half-lives; these intervals are generally wide but for a majority of countries finite, providing an extra layer of support for the PPP hypothesis.[8]

From a practical perspective, half-live estimates of three to five years compare with a typical international portfolio investment horizon of one to two years. Similarly, the magnitude and high volatility of PPP deviations present particular difficulties for risk-averse investors concerned with the volatility and drawdown characteristics of portfolio returns as well as the sign of these returns. As a result, the naive PPP hypothesis has limited applicability in any practical financial context. Another shortcoming of PPP from an investment portfolio context is the issue of endogeneity, and therefore causality. PPP emphasises the tendency of national price levels and the nominal exchange rate to adjust to shocks to ensure either level or trend stationarity of the real exchange rate over time. But neither element of this relationship is exogenously determined, implying that nothing can be stated a priori about the directional transmission of shocks. In an investment portfolio context, the directional causality of shocks to PPP-based equilibria is a crucial issue. Implementing portfolio management decisions on the basis of a theory that embodies two-way causality or, worse, a lead-lag relationship that runs from the exchange rate to prices, is inappropriate. It is important, therefore, that the source and persistence of deviations from PPP-based equilibria are understood and addressed.

By contrast to absolute PPP that we have discussed so far, relative PPP allows for the presence of a unit root in ε_t, and relates changes in the nominal exchange rate to changes in national price levels, that is

$$\Delta s_t = \alpha + \beta(\Delta p_t - \Delta p_t^*) + \varepsilon_t \tag{4}$$

where β is the elasticity of the nominal exchange rate with respect to national prices. Relative PPP, which is implied by absolute PPP[9] holds if one is unable to reject the null hypothesis that $\beta = 1$, even in the presence of a non-stationary error term. The difficulty in testing the validity of relative PPP reflects the need to disentangle the impact on the nominal exchange rate of changes in national price levels from persistent shocks to the error term. Coakley, Flood and Taylor (CFT, 2002) tackle this issue by employing a nonstationary panel estimator using monthly data for various Developed and Developing country exchange rates versus the dollar over the sample period 1970 to 1998. Using this approach, they find widespread evidence in favour of relative PPP. In contrast to conventional wisdom, the results of CFT therefore demonstrate that a cointegrating relationship can exist between a set of variables despite the presence of a non-stationary error process. This finding may have important implications for absolute PPP as well, if the persistence of deviations reflects the existence of real variables that form a cointegrating vector with exchange rates and national price levels

[8] It is interesting to note, however, that the minority group of exchange rates for which the upper confidence interval is infinity includes both the US dollar and yen.

[9] The converse is not true.

that are not incorporated into the traditional absolute PPP relationship. It is to this issue that we now turn.

2.3 Explaining Divergence from PPP-Based Equilibria

A number of reasons have been proposed to explain the persistence of real exchange rate deviations from PPP. Fundamental economic explanations include the incidence of monetary and real shocks to the exchange rate, for instance as propounded by HBS. We discuss the HBS effect in detail in the following section. Other, microeconomic factors include the existence of non-trivial transaction costs (O'Connell and Wei, 1997), sticky price adjustment and the practice of Pricing-to-Market (PTM, Kettner, 1983). PTM involves exporters adjusting the price markup levied in markets that experience a currency depreciation against the exporter's domestic currency and is undertaken in order to stabilise the domestic currency price of its production from a buyer perspective; an alternative interpretation of PTM-due to Krugman (1987)-is that this practice reflects the discriminatory pricing behaviour of profit maximising monopolistically competitive firms (Cheung, Chinn and Fujii, 1999). Knetter finds evidence that PTM is widespread amongst US, Japanese, German and UK exporters, implying that the persistence of deviations from PPP over short time horizons is common across numéraire currencies. In addition, this behaviour may also give rise to non-linear adjustment in the real exchange rate around linear (or log-linear) PPP-based equilibria. In a similar vein, Cheung, Chinn and Fujii (1999) find that a range of structural impediments to market efficiency exert a significant positive impact upon the persistence of PPP deviations. The conclusions of Chen and Devereux (2003) on the basis of price data for nineteen US cities over the period 1918 to 2000 are complementary to the findings of Cheung, Chinn and Fujii, and supportive of PPP: price dispersion has generally fallen over the course of their sample period as market integration between US cities has increased, and price dispersion is lower for traded than non-traded goods. Consequently, any changes to market structure that encourage inter-country trade integration appears likely to reduce the persistence of PPP deviations.

The persistence of exchange rate deviations from PPP may also reflect a number of important statistical issues. These include the low power of conventional unit root tests (Lothian and Taylor, 1996), the relatively short span of exchange rate data used in most empirical studies (including, by necessity, this one) and data measurement error, for instance due to aggregation bias in data samples caused by sector heterogeneity (Taylor, 2000; Imbs, Mumtaz, Ravn and Rey, 2002); Taylor (2000) concludes that a four or five fold upward bias can be introduced into traditional estimates of half-lives by a combination of these factors, reducing half-life estimates to a level consistent with the presence of nominal rigidities alone. Although other researchers question the size of this type of bias, compared with those introduced by the choice of estimator as discussed above (Chen and Engel, 2004), it is widely accepted that their correction will tend to reduce the persistence of real exchange rate deviations from PPP-based equilibria to some extent.

2.3.1 Harrod-Balassa-Samuelson Hypothesis

Starting with Harrod (1933), Balassa (1964), and Samuelson (1964), a large literature has demonstrated that productivity or real incomes levels can systematically influence the relative prices of traded and non-traded goods within a country and hence international relative price levels across countries and time (Dornbusch, 1988). According to HBS relatively fast growth in the traded goods sector of an economy will typically cause an appreciation of the real exchange rate. If correct, this hypothesis implies that an understanding of the source of productivity growth within a country, rather than simply a comparison of aggregate growth rates between countries, is crucial to an understanding of the equilibrium path of the real exchange rate. Accepting the existence of temporary price and wage stickiness also allows for some impact of the HBS effect upon the nominal exchange rate.

To understand the intuition of the HBS hypothesis assume the existence of two economies, home and foreign, that produce a combination of identical tradable and non-tradable goods. Also assume that transaction costs and artificial barriers to cross-border trade and capital flows are zero, that tastes are homothetic and that PPP holds continuously in the tradable goods sector. In addition, assume that labour is homogeneous within each economy and that there are no artificial impediments to inter-sector labour mobility. Finally, assume that prices are set across both economies to equal marginal costs, so that perfect competition applies. All of these assumptions can be relaxed to some degree in practice without fundamentally compromising the conclusions of the HBS hypothesis. But for now they help clarify its implications.

From equation (1) the real exchange rate can be written in logarithms as

$$q_t = s_t + p_t^* - p_t. \tag{5}$$

Under the HBS hypothesis national price indices, p and p^*, can be decomposed into tradable (T) and non-tradable (N) components,

$$p_t = \alpha p_t T + (1-\alpha)\, p_t N \tag{6}$$

$$p_t^* = \beta p_t^* T + (1-\beta)\, p_t^* N \tag{7}$$

where α and β represent the contribution to total value-added of the traded goods sector in the home and foreign economies. Assuming that PPP holds continuously in the traded goods sector gives

$$q_t T = s_t + p_t^* T - p_t T \tag{8}$$

which is the PPP condition (2) above applied only to the traded goods sector. Substituting (6), (7) and (8) into (5) gives the key HBS relationship,

$$q_t = q_t T + (\alpha - 1)(p_t T - p_t N) + (1-\beta)(p_t^* T - p_t^* N). \tag{9}$$

The assumption that absolute PPP holds on a continuous basis in the tradable goods sector implies that $q_t T$ in (9) will be constant, or unity if we assume strict-form PPP.

The HBS transmission mechanism begins with a rise in the level of traded goods sector wages in the home economy in response to an increase in labour productivity. By implication, this improvement in productivity will already have been reflected in rates of return on capital and profit margins in this sector. This shock is assumed not to impact the price of tradable goods, as these are set by world market conditions; as noted above, traded goods sector PPP is assumed to hold continuously.

A central assumption of the HBS hypothesis is that wage rates tend to equalise across the traded and non-traded goods sectors of an economy, reflecting perfect labour homogeneity and mobility. As a result, higher wage levels in the traded goods sector trigger a rise in service sector wages also. In the absence of an offsetting improvement in non-traded goods sector productivity, this wage increase pushes up the average price level in the non-traded goods sector as firms act to maintain prices equal to marginal costs, and hence raises aggregate price levels in the total economy. Assuming that the real exchange begins at a level equal to PPP, this means that the domestic currency becomes overvalued on a naive PPP comparison, because of non-traded goods sector inflation. Consequently, according to the HBS hypothesis a positive productivity shock emanating from the traded goods sector will tend to generate an appreciation of a country's currency, beyond its PPP level.[10]

Researchers have traditionally invoked the HBS hypothesis to explain why the real exchange rate can trade above the level implied by PPP for extended periods. This directional bias stems from the assumption that productivity innovations typically occur within the traded goods sector, with associated wage and price implications subsequently transmitted to the less dynamic non-traded goods sector. Consequently, inter-country productivity differentials in the non-traded goods sector are typically assumed to be much smaller than in the traded goods sector (Balassa, 1964). As productivity differentials in the output of tradable goods grow between countries, difference in wage and price levels of non-tradable goods will also increase. This in turn will lead to a growing wedge between the real exchange rate and PPP. This traditional interpretation also implies that strong growth in aggregate real GDP will be positively correlated with a real appreciation of the home currency, since it is typically assumed to result from technological innovation and productivity catch-up concentrated in the traded goods sector (Balassa, 1964; Samuelson, 1964).

Relatively little consideration has been given to the possibility that the directional bias in the real exchange rate relative to PPP introduced by the HBS effect may be negative, and that strong GDP growth could be associated with a weaker real exchange rate relative to equilibrium. One recent noteworthy exception is Chen and Rogoff (2002), who provide an application of the HBS hypothesis to the Australian dollar consistent with this interpretation. For a similar discussion, see Sager, Nuttall and Taylor (2002). Alternatively, therefore, assume that the shock to productivity emanates from the nontraded goods sec-

[10] Expectations of subsequent improvements in income will reinforce the impact of productivity innovations on the real exchange rate as well, as both consumers and investors smooth future expenditure (Bailey, Millard and Wells, 2001).

tor (Ito, Isard and Symansky, 1997). Productivity gains in this sector translate into non-inflationary wage rises in this sector that are mimicked in the traded goods sector but not backed by corresponding productivity growth. With perfect labour mobility between sectors, higher non-tradable productivity translates into a rise in the price of tradable goods that in turn leads to a depreciation of the real exchange rate as foreign markets tend to switch away from purchasing the higher priced goods.

For completeness, and by implication, as long as the non-traded goods sector is more labour intensive than the traded goods sector, balanced positive productivity shocks to both sectors will have a net effect similar to a positive non-traded goods sector productivity shock and will also weaken the real exchange rate relative to its PPP level (Froot and Rogoff, 1994; Sarno and Taylor, 2002).

The HBS hypothesis has been practically applied to a wide range of Developing (for example, Ito, Isard and Symansky, 1979; Chinn, 1998; Crucini, Telmer and Zachariadis, 2000; Chen and Rogoff, 2002) and Developed (De-Gregorio and Wolf, 1994; Canzoneri, Cumby, Diba and Eudey, 1998) exchange rates, using a variety of statistical methodologies. These studies provide a range of conclusions, such that evidence on the validity of HBS remains inconclusive. None of the studies are particularly definitive, given issues of data coverage and availability. Perhaps most appropriate, as it circumvents aggregation and measurement issues inherent in national price and productivity data, is the use of disaggregated price data. Noteworthy here is the contribution of Crucini, Telmer and Zachariadis (CTZ, 2000), who examine Eurostat data on 3,500 individual tradable and non-tradable price series drawn from European Union countries to examine the source of PPP violations. Consistent with the HBS hypothesis, CTZ conclude that price variation is typically more extensive in the non-traded goods sector, with the assumption of continuous PPP a relatively good approximation of reality for homogeneous tradable goods. But this conclusion is not universally accepted, and Isard (1977) and Richardson (1978) both provide contradictory evidence to CTZ also using disaggregated price data; however, both these studies use much less extensive product ranges and data samples than CTZ. Similarly, Engel (1993) concludes, based upon volatility analysis of disaggregated price data, that the relative price of non-traded goods has little relevance to movements in the US dollar; although it is questionable whether his evidence actually contradicts the HBS hypothesis, as the latter emphasises persistent trends around PPP, rather than the volatility of deviations.

Disaggregated price data across an extensive range of products are rare indeed. Consequently, most researchers are required proxy the HBS effect using aggregate price and productivity data series. One approach-to which we resort in this study-is to use output and employment from the manufacturing sector, maintaining the assumption that service sector productivity growth is constant. Alternatively, some researchers use directly measured data on sector-based estimates of labour and total factor productivity, thereby allowing a role for productivity innovations in the non-traded goods sector; for instance, De-Gregorio and Wolf, 1994; and Canzoneri, Cumby, Diba and Eudey, 1998. But these data are typically only available on an annual frequency, often with a considerable publication lag, implying that any inherent non-linear structure within exchange rate and productivity data may be undermined, and inconsistencies between measurement methodologies across countries also represent a major hurdle.

Alternatively, a number of studies proxy inter-sector productivity differentials using a ratio of consumer (CPI) to producer price (PPI) indices, often finding in favour of the HBS effect (Closterman and Schnatz, 2000; Chinn, 2000; and DeLoach, 2001). But there are a number of pitfalls associated with this approach, including differences in methodologies, weights and baskets used in constructing PPI and CPI indices within and between countries, and the confusion of traded and non-traded goods included in both price indices. As an alternative, the GDP deflator can be decomposed into its tradable and non-tradable components: Ito, Canzoneri and Symansky (1997) adopt this approach to examine the efficacy of the HBS effect for a range of Asian economies over a maximum sample from 1960 to 1990, concluding that evidence in favour of the HBS effect is equivocal. Although these data can achieve a cleaner separation of tradable and non-traded goods sectors within individual countries, methodological differences between countries in the division of output into real and nominal components can make this an inferior measure of the HBS effect (IMF, 2001).

Other studies examine the HBS hypothesis on the basis of total GDP to the number of employed persons, again with mixed results: for instance, Osbat, Rüffer and Schnatz (2003) conclude in favour of a substantial HBS-related productivity effect in a study of euro-yen, whereas Schnatz, Vijselaar and Osbat (2003) conclude that productivity developments had only a marginal role in the depreciation of the euro against the dollar following its introduction in 1999, once data inconsistencies between the Euro Area and the US have been corrected. Other studies proxy the HBS effect using traded and non-traded GDP data divided by the number of persons employed: using this approach Canzoneri, Cumby and Diba (1996) report results favourable to the HBS hypothesis in a panel study of thirteen OECD countries; by contrast, Chinn, 1998 reports mixed results for the HBS effect in an application to East Asian currencies. Chinn (1998) also examines-and rejects-the role of per capita GDP in explaining the persistence of PPP deviations-his premise is that growing per capita income may lead to a rising preference for non-traded goods and services; he is also unable to find a significant role for Terms of Trade shocks in the determination of these real exchange rates.

Little work has yet been undertaken incorporating the assumption of non-linear dynamics into an HBS-augmented PPP model (an exception is the long span study of Lothian and Taylor, 2004). This is something that we address in this study, consistent with the implication of Sarno and Taylor (2002) that stronger and more supportive evidence in favour of the HBS hypothesis may be achieved by its application within a non-linear cointegrating model.

2.3.2 Competing Theories of Real Exchange Rate Determination and Productivity

In addition to the HBS hypothesis, there exist alternative explanations of the relationship between productivity innovations and the real exchange rate. These include the general equilibrium models of Stockman (1980) and Lucas (1982). These models examine the utility maximisation problem of agents in the context of budget and cash-in-advance constraints. For a succinct discussion of the structure and wider implications of these models, see Sarno and Taylor (2002). The implications of productivity innovations for the real exchange rate within the framework of these models is equivocal, and will depend upon the relative importance of two, rival transmission channels. First, a relative price channel, whereby productivity innovations in the home economy generate an increase in domestic output that can only be absorbed by the market via a reduction in price. This implies that higher productivity is consistent with a depreciation of the real exchange rate, and runs counter to the HBS prediction of an appreciation. Second, a money demand channel, whereby higher productivity induces an increase in money demand and an appreciation of the real exchange rate. Clearly, the net impact of productivity innovations on the behaviour of the real exchange rate will depend upon the relative strength of these two channels, and could differ between exchange rates.

2.3.3 Alternative Supply & Demand Shocks to the Real Exchange Rate

Productivity innovations can potentially help to explain the persistence real exchange rate divergence from PPP-based equilibria. But this persistence may also reflect other-supply and demand-shocks that provide a boost to real income and spending. These may include shocks to the Terms of Trade and real government spending (Mussa, 1984; Froot and Rogoff, 1994; Faruqee, 1995; Chinn, 1998; Cheung, Chinn and Fujii, 1999; DeLoach, 2001; and Osbat, Rüffer and Schnatz, 2003). A country's Terms of Trade is defined as the ratio of export to import prices, both expressed in domestic currency terms. There are at least two channels through which Terms of Trade shocks can be transmitted to the real exchange rate. First, due to changes in consumer preferences in favour of the output of the domestic country that raise its Terms of Trade, and appreciate the real exchange rate. And Second, the Terms of Trade of the domestic economy will increase due either to a shift in foreign demand patterns towards its higher value exports, or due to a commodity price shock that favours the production base of the domestic economy. Again, the result will be an appreciation of the real exchange rate.

To the extent that increases in government spending concentrate upon the non-traded goods sector, this will tend to raise the price of non-tradable goods and services and therefore lead to an appreciation in the real exchange rate in the short term, consistent with Dornbusch's (1998) observation that price levels are typically high in borrowing countries.[11] But consistent with the work of Beveridge and Nelson (1991), demand-driven shocks to the real exchange rate are likely to be less persistent than supply based shocks, suggesting that the latter are the more probable source of persistent PPP disequilibria discussed so far in this paper.

The HBS hypothesis can be augmented, at the cost of diluting the assumption of perfect capital mobility, to explicitly include Terms of Trade differentials between home (tt) and foreign (tt^*) countries, and real government spending as a percentage of home (g) and foreign (g^*) GDP.[12] So doing amends equation (9) to

$$q_t = q_tT + (\alpha - 1)(p_tT - p_tN) + (1 - \beta)(p_t^*T - p_t^*N) + \gamma(tt_t - tt_t^*) + \zeta(g_t - g_t^*), \tag{10}$$

where we assume $\delta q/\delta\gamma, \delta q/\delta\zeta > 0$. Adopting this approach, DeGregorio and Wolf (1994) provide empirical support for the hypothesis that both supply and demand-related shocks contribute in a significant manner to an explanation of persistent deviations of the real exchange rate from PPP.[13] Similarly, DeLoach (2001) finds circumstantial evidence supportive of the HBS effect-which he proxies using a ratio of CPI to wholesale prices (WPI)-in a study of nine OECD countries. However, he falls short of demonstrating the existence of a cointegrating equilibrium for any of these nine countries using this approach. Although his findings are strengthened when the basic model specification is augmented to include the real oil price, he still fails to find consistent evidence of cointegrating equilibria.[14] Koen, Boone, de Serres and Fuchs (2001) similarly emphasise the role of the oil price in the behaviour of euro-dollar following its inception in 1999. Although we attempt to include the real oil price within our analysis in a manner consistent with DeLoach's (2001) definition, we were unable to find evidence of a cointegrating vector that incorporated the real exchange rate, productivity differentials and the real oil price; indeed, while Augmented Dickey Fuller (ADF) tests-reported below-suggest that the real oil price is an I(1) variables similar to other series in our study, visual inspection of the data

[11] Adverse credibility effects may reverse this impact in the longer term.

[12] Data availability and comparability issues, as will as the conclusions of Beveridge and Nelson (1991), deterred us from including government spending within our analysis. In addition, although other studies also include long-term interest rate differentials in this regression, we do not, for two reasons: first, becasue we wish to focus upon the validity of the HBS primarily, rather than a general explanation of PPP deviations; and second, because interest rate differentials between Developed countries are typically found to be I(0) series, whereas we conclude that each of our four real exchange rates are I(1) series (Table 1). It is consequently impossible for these variables to form a cointegrating vector.

[13] The finding of a significant and correctly signed relationship between the real exchange rate and government spending contrasts with the findings of other researchers; for instance, although Chinn and Johnston (1997) report an appropriate relationship between US government spending and the real exchange rate, the same is not true for foreign spending.

[14] Similar to the Terms of Trade, the real oil price is intended primarily to capture the impact of supply shocks upon the real exchange rate in addition to productivity differentials. The sign of the estimated relationship between the real oil price and exchange rate more equivocal than for the Terms of Trade, for instance depending upon the import dependency of a country.

casts some doubt upon this conclusion. Furthermore, our failure to discover an appropriate cointegrating vector that includes the real oil price is consistent with the results of Alquist and Chinn (2000). But we do find evidence in favour of incorporating the Terms of Trade, as well as productivity differentials, in our empirical modelling, as is discussed below.

2.3.4 Smooth Transition AutoRegression Models

The discussion of the previous section suggested that inclusion in exchange rate models of supply related real variables–and particularly relative productivity differentials and the Terms of Trade–can help diminish the persistence of shocks to the real exchange rate around PPP-based equilibria. In this section, we examine whether a further reduction in persistence may be achieved by modelling the equilibrium correction mechanism that exists between the real exchange rate, productivity differentials and the Terms of Trade as a non-linear process.

A recent, but growing, literature has found in favour of the existence of significant neglected nonlinearity in exchange rate dynamics (Michael, Nobay and Peel, 1997; Taylor and Peel, 2000; Kilian and Taylor, 2001; Taylor, Peel and Sarno, 2001; Taylor and Sarno, 2001). A common approach amongst early non-linear research was to model exchange rate dynamics within a Threshold AutoRegressive (TAR) framework (Tong, 1983, 1990). In this case, the existence of non-trivial arbitrage costs, product differentiation and uncertainty as to the exact value of equilibrium, introduces a threshold on either side of Fair Value, for instance, up to one standard deviation, within which deviations from Fair Value are relatively persistent. Consequently, within these thresholds the real exchange rate exhibits unit root, or even explosive, behaviour. As the real exchange rate reaches these thresholds, consensus emerges amongst investors that it has moved significantly away from equilibrium, implying that the benefit of arbitraging this opportunity now outweighs the fixed cost. Consequently, the exchange rate will tend to revert back towards its equilibrium level beginning around the level of these thresholds. TAR models therefore assume that the exchange rate shifts abruptly between unit root and mean reversion states, with this shift occurring around the level of the threshold.

The general form of a TAR model can be written as

$$\Delta q_t = \alpha + I_{1t}(q_{t-d} - m) + (1 - I_{1t})(q_{t-d} - m) + \varepsilon_t \tag{11}$$

where m is the estimated equilibrium level of the real exchange rate (q_t); q_1, q_2 are threshold parameters, ϵ_t is a residual term assumed to be independently and identically distributed with variance σ, and

$$I_{1t} = \left\{ \begin{array}{c} 1 \text{ if } q_{t-d} - m \geqq q_1 \\ 0 \text{ otherwise} \end{array} \right\} \qquad 1 - I_{1t} = \left\{ \begin{array}{c} 1 \text{ if } q_{t-d} - m \geqq q_2 \\ 0 \text{ otherwise} \end{array} \right\}. \tag{12}$$

That the behaviour of the exchange rate shifts abruptly between two behavioural states in proximity to a specific threshold level appears to be an overly restrictive assumption. More plausible is that shifts between these two states occur smoothly, so that the tendency to revert back towards Fair Value is a positive function of the size of disequilibrium (O'Connell and Wei, 1997; van Dijk and Franses, 2000; Peel, Taylor and Sarno, 2001; Sarno and Taylor, 2002). This assumption is consistent with the observed presence within the foreign exchange market of heterogeneous investors, and therefore with the observation that the perceived magnitude of realisable profits from arbitrage at any time will vary between market participants, for instance due to differences in the transaction costs that each investor group faces and the speed with which each group learns of the disequilibrium and the existence of a significant arbitrage opportunity.[15] It also appears consistent with the view of Sarno and Taylor (2002) that the probability of successful central bank foreign exchange intervention increases with the size of disequilibria, with the central bank ultimately playing a coordinating role to the arbitrage activity of other market participants-who individually may be uncertain as to the extent of any disequilibrium or of the probability that market participants collectively will arbitrage available opportunities-to drive the exchange rate back towards its equilibrium level.

Smooth Transition AutoRegressive (STAR) models capture this type of nonlinear adjustment dynamics. This class of models was popularised by Granger and Teräsvirta (1993), and has subsequently been widely applied in macroeconomic and financial empirical modelling. For a comprehensive discussion of the STAR methodology, as well as non-linear modelling more generally, see Granger and Teräsvirta (1993), Franses and van Dijk (2000), and van Dijk, Franses and Teräsvirta (2002).

The general form of a STAR model may be written as

$$\Delta q_t = \alpha + \beta \Delta x_t + f\{z_{t-d}\}(\alpha' + \beta' \Delta x_t) + \varepsilon_t, \quad (\beta + \beta') < 1 \tag{13}$$

where Δq_t is the real exchange rate as defined in equation (5) above, and included in first differences to ensure stationarity (a prerequisite of STAR modelling). X_t is a set of explanatory variables, in our case productivity differentials, the Terms of Trade and lagged values of Δq_t. $f(z_{t-d})$ is a bounded continuous function in z_{t-d} that characterises the transition of q_t between the two regimes embedded within STAR models. The transition variable z_t is usually chosen on the basis of theoretical intuition. For many financial and economic series there would appear to exist a number of series that represent plausible transition variables. The range of possibilities is more limited for exchange rate modelling, and is typically limited either to the dependent variable itself–in which case the model is termed a Self-Exciting AutoRegressive or SETAR model–or to the equilibrium correction term that results from the underlying linear cointegrating vector (Taylor and Peel, 2000; Sarno and Taylor, 2001). Again, a crucial determinant in the choice of transition variable is that the series be stationary.

[15] For a comprehensive discussion of foreign exchange investor heterogeneity, see the chapter in this study on The Role of Order Flow in Exchange Rate Forecasting.

There are two general types of STAR model: exponential and logistic. For an exponential STAR (ESTAR) the non-linear transition function in equation (13) takes the form

$$f(z_{t-d}) = 1 - exp\{-(\gamma/(\sigma(q)))((q_{t-d} - c)^{\hat{}2})\}. \tag{14}$$

The location parameter c determines the level at which the transition occurs between the two regimes embedded in the model. The parameter γ is often termed the smoothness parameter (Enders, 2003) and governs the speed of transition between the unit root and mean reversion states embedded in the model. It takes a value $0 \leq \gamma \leq \infty$. High values of γ imply that the underlying adjustment process is similar to the TAR model in equations (11) and (12) above, where shifts between the two states occurs instantaneously at some threshold level (Teräsvirta and Anderson, 1992); indeed, the two models are equivalent for $\gamma = 1$. By implication, estimated values of γ close to zero indicate that the transition between states occurs very slowly. The term $\sigma(q_t)$ is the standard deviation of the transition variable. It is included in the transition function as a scaling variable to facilitate convergence of γ to its true value during the estimation procedure-this is sometimes problematic, particularly in small samples-and to allow comparison of the speed of transition between unit root and mean reversion regimes across each of our four exchange rate models.

The ESTAR model is symmetric around $f(z_{t-d})$. Consequently, $f(zt - d)$ is U-shaped, and the rate at which the real exchange rate reverts back to equilibrium is equivalent for large positive and negative disequilibria of similar size. When the exchange rate is trading at or close to its equilibrium level, that is when $q_{t-d} - c$ is approximately zero, the ESTAR transition function collapses to zero so that Δq_t will be a function of the underlying linear AR process,

$$\Delta q_t = \alpha + \Sigma\beta_i \Delta x_{t-i} + \varepsilon_t. \tag{15}$$

For large deviations of the real exchange rate from equilibrium, $f(z_{t-d}) = 1$ and Δq_t becomes a function of an alternative, non-linear AR process given in equation (13) above,

In many studies neglected non-linearity is assumed to be present in all of the behaviour of the dependent variable, Δq_t, consistent with estimation of equation (13). By contrast, in this study we assume that non-linearity is confined to the adjustment of the real exchange rate around its underlying (linear) equilibrium path. Our approach is consistent with the approach of Hendry and Ericsson (1991) in a study of UK broad money demand who, to the present authors' best knowledge, were the first to restrict interaction between the non-linear function and the linear AR representation of a series to the equilibrium correction term. Their approach was relatively simple, and involved multiplying the standard equilibrium correction term by its own squared value. By contrast to this study, there was no attempt to model the speed of adjustment of broad money around its equilibrium value as a function of the magnitude of the divergence of the two series; instead convergence occurred at a constant rate. This decision leads us to amend (13) to,

$$\Delta q_t = \alpha + \beta \Delta x_t + f\left(z_{t-d}\right) \phi \Delta x_t + \varepsilon_t. \tag{16}$$

where

$$f(z_{t-d}) = 1 - exp\{-(\gamma/(\sigma(ECM)))(EC\hat{M}_{t-d}^{2})\}. \tag{17}$$

Lag length, d, of the transition variable is determined on the basis of LM-type linearity tests proposed by Granger and Teräsvirta (1993) and Teräsvirta (1998), where d is chosen to maximise the rejection of the null hypothesis of linearity, HO: $\gamma = 0$, in favour of the null of a STAR-based non-linearity. We discuss these tests in more detail below.

We adopt the ESTAR specification as our central case. This follows from the fact that although we will set out to demonstrate the presence of some form of non-linear adjustment process in the real exchange rate around its equilibrium path, there is no presupposition that the speed of this adjustment is different in periods of significant over- and under-valuation. Asymmetry may be present for other financial variables–for instance, equity prices, where short-selling restrictions may inhibit the activities of some investors–or real economic variables, where periods of sub-trend growth are typically shorter than above-trend periods; for instance, see Teräsvirta and Anderson (1992) for an application of STAR modelling to Developed country industrial production data. By contrast, it is difficult to conceive of intuitive reasons why Developed country exchange rates should exhibit similar asymmetries. One potential justification may be the intervention activity of central banks that could exhibit a bias to prevent either over- or under-valuation. But outside of the Bank of Japan, it is not clear that this directional bias forms part of the objective function of any major central bank. And in any case the extent of official intervention has diminished substantially in recent years.

Nonetheless, our choice of an ESTAR specification is challenged by the findings of Enders and Dibooglu (2001) and Leon and Najarian (2003). Using threshold cointegration, Enders and Dibooglu report evidence of asymmetric adjustment around PPP-based equilibrium for seven European exchange rates using monthly data over the sample period 1973 to 1997. In this case, a logistic transition function would be more appropriate, turning the model into an LSTAR. Similarly, Leon and Najarian find evidence of asymmetric adjustment in an examination of monthly real effective exchange rate data for twenty six Developed and Developing country exchange rates using a Bi-Parameter Smooth Transition AutoRegressive (BSTAR) model; this is a variant on the STAR framework discussed above that incorporates two-speed convergence to reflects the existence of investor heterogeneity in the foreign exchange market. This evidence of asymmetry is strongest for Developing country exchange rates during episodes of over-appreciation.

For completeness, therefore, we also model the adjustment of the real exchange rate around equilibrium as an LSTAR process that incorporates the notion of asymmetric adjustment and compare the results with the ESTAR formulation. In the LSTAR formulation the transition function $f(z_{t-d})$ takes the form

$$f(z_{t-d}) = (1 + exp\{-(\gamma/(\sigma(ECM)))(ECM_{t-d})\})^{-1}, 0 \leq \ \gamma \leq \ \infty.$$

In contrast to the ESTAR specification the transition function in an LSTAR model is a monotonically increasing function of ECM_t in our application, or of z_{t-d} more generally.

Specifying a STAR Model Although only capable of capturing one of many potential types of non-linear behaviour STAR models have proved popular in the academic literature partly because there exists a straightforward framework, established by Granger and Teräsvirta (1993), Teräsvirta (1994) and Teräsvirta (1998), for specifying and testing the performance of this group of models.[16] This framework incorporates three steps:

1). specify a parsimonious linear equilibrium correction model that incorporates a cointegrating equilibrium between the dependent and explanatory variables. This linear VECM should exhibit no evidence of residual non-normality, autocorrelation or heteroskedasticity, or parameter instability;

2). test this optimal linear specification for evidence of neglected non-linearity, across a range of plausible values of the delay parameter, d. This test amounts to a test of $Ho: \gamma = 0$ in equations (13) and (16) above;

3). if linearity is rejected, choose the appropriate form of STAR model, evaluated against a range of misspecification tests, as detailed below.

A parsimonious linear VECM model may be written as

$$\Delta y_t = \alpha + \beta_i \sum_{i=1}^{p} \Delta x_{t-i} + \varphi(y_{t-1} - x_{t-1}) + \varepsilon_t \tag{18}$$

where $\varphi(y_{t-1} \ - \ x_{t-1})$ is the linear equilibrium correction term and ε_t is an error term assumed independently and identically distributed with mean zero and variance σ. We select the optimal lag length for each VECM using a Wald Likelihood Ratio test; this test assesses the joint significance of all i-th lagged endogenous variables in the VECM, and has a chi-square distribution with k^2 degrees of freedom under the exclusion null. VECMs are estimated using FIML and optimised using a traditional general-to-specific procedure, with the least significant coefficient removed from the system at every iteration until all coefficients are significant at a 5% level.

In this linear framework the dynamic relationship between the real exchange rate and significant regressors is assumed to be constant through time, meaning that the real exchange rate adjusts back towards its Fair Value in the wake of a shock at a constant rate φ in all periods independent of the magnitude of the deviation from equilibrium. This is a key difference with the STAR models.

[16]STAR models can also capture the effect of non-constant parameters within the linear alternative.

Assuming optimised VECMs pass all diagnostic tests, we progress to the second stage of the Granger-Teräsvirta modelling process. This involves testing the optimal linear specification for evidence of neglected non-linearity of the STAR form. As Teräsvirta and Anderson (1992) discuss, the purpose of linearity testing is twofold. First, to isolate those exchange rates for which the null hypothesis of linearity cannot be rejected, thereby eliminating these from the modelling process and, second, to determine an estimate of the delay parameter for those exchange rates for which non-linearity is present. The most common approach to linearity testing is to apply a variant of the Lagrange Multiplier linearity test proposed by Granger and Teräsvirta (1993) and Teräsvirta (1998). This test involves a Taylor expansion of a general STAR model around powers of the chosen transition variable, resulting in the following auxiliary regression,

$$\varepsilon_t = \beta_0 x_t + \beta_j \sum_{j=1}^{p} x_{t-j} ECM_{t-d}^{j} + u_t, \; H0 : \beta_j = 0, \quad j = 1, 2, 3 \qquad (19)$$

where ε_t is the residual series from the optimal linear VECM in (14), $\sum_{j=1}^{p} x_{t-j}$ is the set of regressors from this VECM and ECM_{t-d} is the equilibrium correction term, our chosen non-linear transition variable. In small samples the test is distributed as an F-test with $(m - 1/N - m)$ degrees of freedom, where m is the number of regressors in the auxiliary regression and N the number of sample observations. The test is calculated by comparing the sum of squared residuals from this auxiliary regression with the same statistic from the optimal linear VECM. In order to determine the optimal value of d, the delay parameter, the test is calculated over a range of values, $1 \leqslant d \leqslant D$, where the value of D is determined by economic intuition. A significant rejection of the null hypothesis for any value of d is interpreted as indicative of the presence of a non-linear STAR process within the residuals of the estimated linear VECM; equivalently, the conditional mean of the dependent series is non-linear. If two or more values of d are significant, the optimal value of d, $\hat{d}$, is chosen to be the one that minimises the p-value of the test statistic. As an incorrect rejection of the null hypothesis of this linearity test will be revealed by the subsequent failure to a uncover a STAR model that provides a satisfactory explanation of the data (Teräsvirta and Anderson, 1992), the researcher can afford to adopt a relatively liberal attitude to acceptable significance levels when conducting this test.

An alternative test often employed to confirm the presence of neglected non-linearity is the Ramsey Reset test (Ramsey, 1969). Although originally designed as a test for evidence of general misspecification, a significant result is typically interpreted as evidence of either non-linearity or time-varying parameter estimates, the impact of which can also be captured by the STAR methodology. The Ramsey Reset test is based upon the following regression,

$$\Delta y_t = \alpha + \beta_i \Delta x_{t-i} + \zeta_i Fitted_t\,\hat{}^{\,j} + \varepsilon_t \qquad (20)$$

where $\zeta_i Fitted_t\hat{}^j$ are powers of the fitted values from the basic regression of Δy_t on $(\alpha + \beta_i \Sigma_{i=1}^{p} \Delta x_{t-i})$ in equation (19) above. Similarly, Ljung-Box Q-statistics calculated from the autocorrelation function of the squared residuals recovered from the optimal linear VECM can be used to indicate the presence of some form of misspecification. Researchers also often test for the presence of neglected non-linearity using the BDS test proposed by Brock, Dechert, Scheinkman and LeBaron (1996). Under the BDS test, and following Enders (2003), let $\{\varepsilon_t\}$ be the sequence of residuals recovered from the optimal VECM and τ represents a given distance. If all values of $\{\varepsilon_t\}$ are independent, then the probability that the distance between any two observations $\{\varepsilon_i, \varepsilon_j\}$ is less than τ should be equivalent under the null hypothesis for all i and j, meaning that the residuals are independently and identically distributed. If this is not the case, then there exists misspecification in the residuals, which is often assumed to be some, unspecified form of nonlinearity.[17] As the BDS test has low power in small samples we report p-values calculated from 10,000 iterations of a bootstrap simulation that assumes the underlying series is distributed as a random walk.

Assuming that the null hypothesis of linearity is rejected in equation (20), Granger and Teräsvirta (1993) and Teräsvirta (1994, 1998) suggest using a series of nested F-tests based upon this auxiliary regression to facilitate differentiation between an ESTAR and LSTAR specification. These nested tests incorporate the following null hypotheses,

$$Ho_3 : \beta_3 = 0 \tag{21}$$

$$Ho_2 : \beta_2 = 0 \mid \beta_3 = 0 \tag{22}$$

$$Ho_1 : \beta_1 = 0 \mid \beta_2 = \beta_3 = 0. \tag{23}$$

An acceptance of (22) and rejection of (23) is taken to be indicative of an ESTAR model. Alternatively, a significant rejection of (24),and acceptance of (22) and (23) suggests an LSTAR specification. Teräsvirta (1994) demonstrates, on the basis of a Monte Carlo simulation, that this model selection procedure generally works well both in terms of determining the presence of non-linearity and differentiating between ESTAR and LSTAR formulations. But as van Dijk and Franses (2000) argue, this sequence of nested tested tests is not guaranteed to accurately reveal the correct form of STAR model appropriate to the dependent series at hand; after all, it is based only upon a Taylor expansion approximation of the underlying STAR model. This suggests that the researcher is best advised to fit both forms of STAR model to the data set and use the results of this estimation procedure alongside the above F-tests to determine the appropriate STAR representation.

Third, evidence of neglected nonlinearity in the optimal linear VECM justifies the specification of an appropriate STAR model.[18] Once specified, there

[17] As an alternative to nonlinearity, a significant BDS statistic can be indicative of a chaotic process.

[18] By implication, of course, an insignificant set of results suggests that the linear characterisation of the dependent variable is appropriate and that modelling efforts should terminate at this stage.

exist three evaluation tests designed to assess the specification of STAR models. These tests were developed by Eitrheim and Teräsvirta (1996) and are based upon Taylor expansions of the underlying STAR models.[19] As Eitrheim and Teräsvirta (1996) demonstrate, all of these tests have appropriate small sample properties. The first test examines the optimal STAR model for the presence of residual autocorrelation. As the asymptotic distribution of this test is unknown (Eitrheim and Teräsvirta, 1996), the authors develop an appropriate LM test that compares the sum of squared residuals from the optimal STAR model and an auxiliary regression of the residuals from the optimal STAR model expressed as a function of lagged values of the residuals, the explanatory variables of the STAR model and the two following expressions,

$$\partial f/\partial\gamma = \exp\left\{-\hat{\gamma}(ECM)^2\right\}(y_{t-d}-\hat{c})^2\hat{\theta}'w_t \tag{24}$$

$$\partial f/\partial c = 2\hat{\gamma}\exp\left\{-\hat{\gamma}(ECM)^2\right\}(y_{t-d}-\hat{c})\hat{\theta}'w_t \tag{25}$$

where $f(.)$ is the original ESTAR function. The resulting test is distributed as an F-test in small samples, with q and t-n-q degrees of freedom.

The second test developed by Eitrheim and Teräsvirta examines preferred STAR models for evidence of additional non-linear structure within the real exchange rate. Again, this test is based upon a Taylor expansion of the STAR model and takes the form

$$\Delta\varepsilon_t = \beta_0 x_t + (\theta x_t)\,F_t(y_{t-d};\gamma,c) + \beta_1 x_t y_{t-d} + \beta_2 x_t y_{t-d}^2 + \beta_3 x_t y_{t-d}^3 + \varepsilon_{t,} \tag{26}$$

where $H0 : \beta_1 = \beta_2 = \beta_3 = 0$ and ε_t is the estimated residuals from the optimal ESTAR model. The test is calibrated for values of $d = 1, 2, ..., 8$.

Third, Eitrheim and Teräsvirta (1996) develop a test of parameter constancy against an alternative hypothesis of smoothly changing parameters. This alternative hypothesis contrasts with the traditional assumption of a single structural break in estimated coefficients, for instance as embedded in the Chow test; the alternative hypothesis in the Chow test is a special case of the test proposed by Eitrheim and Teräsvirta. Their test is based upon the alternative STAR model,

$$\Delta q_t = \alpha + \beta(t)\Delta x_t + f\left(z_{t-d}\right)\phi(t)\Delta z_t + \varepsilon_t. \tag{27}$$

To test the null hypothesis of parameter constancy, Eitrheim and Teräsvirta derive the following auxiliary regression,

$$\Delta\varepsilon_t = \beta_i \sum_{i=1}^{3} t^i x_t F_t(y_{t-d};\gamma,c) + \beta_j \sum_{j=0}^{3} t^j x_t + r_{t,}^* \tag{28}$$

where $H0 : \beta_j = 0$. If any of these three evaluation tests indicates the presence of misspecification the selected STAR model should be respecified and only when the researcher is confident that the preferred STAR specification adequately characterises the underlying data generating process of the dependent series should the modelling process terminate.

[19] For a full derivation of these tests, see Eitrheim and Teräsvirta (1996).

2.4 Data

Using quarterly data from DataStream over the maximum sample period 1970:1 to 2002:4 we examine the relationship between four real exchange rates-mark-dollar, yen-dollar, mark-sterling and yen-sterling, all expressed as domestic price of foreign currency-and fundamental variables in a linear and non-linear framework.[20] The real exchange rate is defined as above, in terms of consumer price (CPI) data. We were unable to uncover an appropriate variable that could proxy for the HBS effect across all four exchange rates. Consequently, we use a number of different measures. For mark-dollar, mark-sterling and yen-sterling, traded goods sector productivity is measured as manufacturing output per person employed. According to our interpretation of the HBS hypothesis, an increase in manufacturing productivity in country i is consistent with an appreciation in the real value of its currency. Implicitly, therefore, and consistent with Balassa (1964), for these three exchange rates we assume that non-traded goods sector productivity is constant. For yen-dollar, productivity is also defined as the ratio of manufacturing output per person employed to whole economy output person employed, thereby allowing non-traded goods sector productivity to vary over time for this exchange rate. The Terms of Trade for all exchange rates is measured as a ratio of export to import prices, in domestic currency terms. All fundamental data series are seasonally adjusted, and all series, including exchange rates, are expressed in natural logarithms.

2.5 Empirical Results

Consistent with the preceding discussion, we adopt a four-stage modelling procedure. First, we examine our data set for evidence of unit roots and second,.for evidence of long-term linear equilibria between selected series, using the usual Johansen cointegration procedure. Third, we fit the best linear VECMs around these linear equilibria, and fourth, we test for the presence of non-linear adjustment around these established linear equilibria and then compare the performance of the optimal non-linear model with the optimised linear VECM for each exchange rate. This approach is consistent with Escribano and Miri (1996), Balke and Fomby (1997) and van Dijck and Franses (2000), all of whom demonstrate that the bias in estimating cointegrating relationships between series in the presence of neglected non-linear adjustment is not significantly larger than in the case of linear adjustment. Consequently their results indicate that the presence of non-linear adjustment does not invalidate the use of linear cointegration tests.

[20] The sample for yen-sterling begins in 1984:4, reflecting data availability. Mark-dollar and mark-sterling are used as proxies for euro-dollar and euro-sterling from January 1999. Studies demonstrate that this is approximately equivalent to calculting synthetic euro exchange rates for the period prior to January 1999 (Schnatz, Vijselaar and Osbat, 2003).

2.5.1 Unit Root and Cointegration Tests

A standard first step in empirical research is to examine data series for evidence of a unit root, and then to extend this analysis into an examination of the presence of cointegrating equilibria between variables. This is important to avoid the problem of spurious regressions that occurs when two or more series with no underlying long-term relationship are combined in regression analysis, with some behavioural inference drawn as a result. A spurious regression typically exhibits a high R^2 and significant t-statistics, but these statistics are meaningless in economic terms.

For most time series, and particularly financial market variables including exchange rates, qualitative inspection highlights that the assumption that series are normally distributed with a constant mean, m, and finite variance σ is false. Series that contradict this assumption are termed nonstationary, and have a non-constant mean and a variance that increases over time, to infinity in the limit. A non-stationary series is thereby said to exhibit one or more unit roots, meaning that the impact of a shock to the variable at time t is persistent and does not decay. By contrast, shocks to stationary variables are necessarily temporary and will dissipate over time, with the series ultimately reverting to its long-run mean value (Enders, 1995). Consider the following two series

$$y_t = \alpha + y_{t-1} + \varepsilon_t \tag{29}$$

$$z_t = \alpha + \gamma z_{t-1} + \varepsilon_t, \quad where \;\; 0 < \gamma < 1 \tag{30}$$

y_t is a unit root process and z_t a stationary series. To impose stationarity on y_t, it is necessary to take differences of the original time series. A variable that contains d unit roots and therefore requires differencing d times to impose stationary is termed integrated of order d, or $I(d)$. A variable that is stationary in first differences is therefore termed $I(1)$. As discussed above, an understanding of the order of integration of variables is central to an assessment of the conclusions of associated regression analysis. A necessary but not sufficient condition for two or more variables to form a long-term, linear equilibrium relationship is that they exhibit the same order of integration. By definition, two variables for which the order of integration is different cannot form a long-term relationship, as they will drift apart over time; unanticipated shocks to the estimated relationship will be permanent and there will be no tendency for the previous correlation to re-emerge.

We test variables for order of integration using the Augmented Dickey-Fuller (ADF) test. This test is modified relative to the standard ADF test to incorporate a constant term, and a constant term with drift. Although there exists a plethora of different unit root tests, each with a slightly different slant on this issue, we feel that analysis of these modified ADF tests alone remains appropriate for our purposes. The test takes the form

$$\Delta y_t = \alpha + \zeta y_{t-1} + \sum_{i=1}^{n} \beta_i \Delta y_{t-i} + \delta_t + \varepsilon_t \tag{31}$$

where δ is a time trend and where the error term ε_t is assumed to be independently and identically distributed. Monte Carlo simulations suggest that the significance of ADF test statistics will be biased by the choice of an inappropriate lag structure, leading to the persistence of serial correlation in the error process. Accordingly, we test variables for presence of a unit root structure using a lag length from zero to twenty, with the optimal lag selected on the basis of the Akaike Information Criterion (AIC) to ensure that all serial correlation from the residuals is eliminated.

ADF unit root test results, including optimal lag length, are reported in Table 1. These indicate the presence of a single unit root in most series under examination, suggesting that the real exchange rate and its fundamental determinants are non-stationary in levels. Evidence of a unit root in the real exchange rate is a moot point in the literature. Many studies report findings consistent with our results using similar data samples; for instance, Abauf and Jorion (1990) on the basis of panel unit root tests, and Clarida and Taylor (1997) in a study of the relationship between spot and forward exchange rates for the G5 countries. However, Taylor, Peel and Sarno (2001) using Monte Carlo simulations suggest that standard unit root tests have low power to reject the null hypothesis when the true process incorporates non-linear mean-reversion. As a central hypothesis of our analysis is the presence of nonlinearity in exchange rate behaviour, their conclusion is potentially an important one. O'Connell and Wei (1997) make a similar point, although they also conclude in favour of a unit root using exchange rate data from the floating rate period. Other studies that find in favour of stationarity typically enjoy the benefit of much longer data samples than ours, in some cases several hundred years (for instance, Lothian and Taylor, 1996). This approach, too, can bring its own potential drawbacks, though, as the data span a mix of fixed and floating exchange rate arrangements and, potentially, a number of structural breaks that may have exerted a significant impact upon the behaviour of the real exchange rate.

Three of our selected series appear to be stationary in levels: long-term interest rate differentials, relative government spending-where data is available to allow testing-and the US Terms of Trade. The interest rate and government spending results make intuitive sense. For both variables it seems implausible, in a Developed country setting, that persistent divergence could occur over the whole of our sample without evidence of correction back towards mean levels: interest rate divergence offers riskless arbitrage opportunities to investors beyond some initial threshold that incorporates transaction costs and risk premia; and a widening government deficit in a particular country will not be funded by the market indefinitely, implying eventual correction of relative balances towards a long-term mean.

Variables that are integrated of the same order of integration may form a long-term linear equilibrium, or cointegrating, relationship (Johansen, 1988). Consider a kth-order VECM,

$$\Delta y_t = \nu_t + \sum_{i=1}^{k} \Gamma_i y_{t-i} + \Pi y_{t-k} + \varepsilon_t \tag{32}$$

where Πy_{t-k} is a traditional equilibrium correction mechanism and k is the number of endogenous variables within the VECM system. The Johansen cointegration test examines evidence of linear dependence within the Π matrix; if Π exhibits reduced rank, such that $0 < r < k$, then there exist two r x k matrices α and β, where

$$\Pi = \alpha\beta', \quad \beta' y_t \sim I(0). \tag{33}$$

The rank of Π equals the number of cointegrating vectors, given by each column of the β matrix. The elements of α are the equilibrium correction parameters in the VECM system that determine the speed at which variables return to equilibrium in the wake of an unanticipated shock. Johansen estimates the Π matrix within an unrestricted VAR and tests whether the restrictions implied by the reduced rank of Π can be rejected. As with unit root analysis, one potential problem undertaking cointegration analysis is the relatively short data span available to this study. This implies that the power of cointegration tests may be compromised, reflecting the long-term nature of the embedded hypothesis. With a short data span, regardless of the frequency of observations, it can be difficult to distinguish a mean-reverting series with high persistence from a random walk, and therefore to reject the null hypothesis of no cointegration even if in reality there does exist some linear combination that forms a cointegrating vector. This proviso notwithstanding, we think that this analysis is useful to undertake; indeed, to the extent that we uncover evidence of significant cointegrating equilibria between our selected series then this observation can be taken to strengthen these findings. And although our sample spans a range of monetary and exchange rate arrangements in each of the four countries, not least the UK, that may also compromise the power of these tests, this is an unavoidable feature of many strands of exchange rate modelling[21], and not a characteristic unique to this study.

Table 2 presents the results of cointegration analysis. Optimal lag length for these tests is chosen on the basis of lag exclusion tests and reported in the table, as are assumptions relating to the inclusion of linear time trends within cointegrating vectors. For each of the four exchange rates, the Johansen test indicates the existence of at least one cointegrating vector between the real exchange rate, relative productivity and the Terms of Trade, although the exact composition of these vectors differs between exchange rates, in terms of both the way we include productivity and Terms of Trade variables. The sign of estimated cointegrating coefficients are instructive (Table 3), and suggest that only the productivity terms for mark-dollar are consistent with the presence of a traditional HBS relationship in deviations of the real exchange rate from PPP levels; this finding is consistent with the evidence of Alquist and Chinn (2002),

[21] Use of Markov-Switching models that capture regime shifts can address these issues.

Schnatz, Vijselaar and Osbat (2003) and Osbat, Rüffer and Schnatz (2003) who all argue that relatively weak productivity growth in the Euro Area contributed-to varying degrees-to the depreciation of euro-dollar following its inception in 1999. By contrast, estimated coefficient signs for yen-dollar, yen-sterling and mark-sterling all suggest that positive productivity innovations are consistent with a depreciation of the real exchange rate. Although a contradiction of the HBS hypothesis, this finding is consistent with the existence of a price channel transmission mechanism in the theoretical models of Stockman (1980) and Lucas (1982), whereby output growth implied by positive productivity innovations can only be absorbed by the market if price levels decline. The existence of this transmission mechanism is also consistent with the conclusions of Chinn and Johnston (1997) based upon time series analysis-their conclusions are more consistent with HBS using panel estimation-and IMF (2002); IMF (2002) examines the behaviour of mark-sterling using quarterly data over the period 1980 to 2000, concluding that relatively weak UK productivity has been a significant determinant of the estimated equilibrium revaluation of this exchange rate during the late 1990s.

By contrast, the sign of estimated coefficients for the Terms of Trade series for all exchange rates except mark-sterling is consistent with positive supply shocks other than productivity causing an appreciation of the real exchange rate.

In conclusion, it appears that our results have confirmed the general findings of the extant exchange rate literature: evidence in favour of the HBS hypothesis is mixed, with conclusions differing depending upon selected methodologies, exchange rates and data definitions. But the fact that cointegrating equilibria incorporating relative productivity and the real exchange rate are evident for all of our four rates suggests that we should retain productivity series within our analysis and that supply shocks captured by both productivity and the Terms of Trade do at least in part explain the persistence of PPP deviations. We proceed on this basis.

Consistent with the Granger Representation theorem (Granger, 1993), for any set of variables that form a cointegrating equilibrium there exists an associated dynamic, equilibrium correction representation. Accordingly, Table 4 reports optimised linear VECMs for each exchange rate, expressing quarterly changes in the real exchange as a function of significant lagged changes in productivity differentials, the Terms of Trade, as well as the lagged cointegrating, or equilibrium correction, relationship between all these variables. A significant negative coefficient on this last term ensures that variables within the VECM adjust back towards the long-term cointegrating equilibrium relationship in the wake of an unanticipated shock. We select the optimal lag length for each VECM system using a Wald Likelihood Ratio test; this test assesses the joint significance of all i-th lagged endogenous variables in the VECM, and has a chi-square distribution with k^2 degrees of freedom under the exclusion null. We then use a traditional general-to-specific procedure to sequentially eliminate insignificant coefficients within the system until we achieve a set of parsimonious VECMs within which all remaining coefficients-including equilibrium correction terms-are significant at a 5% level. Exchange rate systems are estimated using the Full Information Maximum Likelihood estimator (FIML). This procedure estimates the likelihood function under the assumption that the contemporaneous errors within the system have a joint normal distribution. Provided that the likelihood function is correctly specified, FIML is fully efficient.

From Table 4, optimised linear VECMs are relatively parsimonious, with few dynamic lagged explanatory terms included in the final specification. The explanatory power of these VECMs is relatively high for all four exchange rates compared with traditional PPP equations; R^2 statistics lie in the range 0.12-for yen-sterling- to 0.24-for mark-dollar. Equilibrium correction terms are correctly signed for all four exchange rates; the speed of adjustment of the real exchange rate back towards equilibrium in the wake of an unanticipated shock is relatively rapid for all exchange rates compared with traditional PPP half-life estimates. The payoff from augmenting PPP with a measure of productivity and the Terms of Trade appears unequivocal.

Diagnostic tests on these optimised VECMs are generally satisfactory, except for evidence of residual non-normality for mark-dollar and a rejection for yen-dollar of the null hypothesis in the Ramsey Reset test of no equation misspecification. Similarly, the autocorrelations of the squared residuals for yen-dollar (Table 5) suggest problems with the optimal VECM for this exchange rate. Some supportive evidence is provided by BDS tests as well, particularly for mark-dollar (Table 6).

Table 7 reports the results of Granger-Teräsvirta (1996) linearity tests on the residuals from the optimal VECM of each exchange rate. These indicate evidence of non-linearity for all exchange rates except mark-dollar. Alongside the Jarque-Bera test result for this exchange rate reported in Table 5, our findings seem consistent with other studies that have demonstrated the presence of non-linear adjustment around (log) linear equilibria for a range of real exchange rates, indicating the importance of modelling exchange rates as non-linear process; for instance, see Taylor and Peel (2000), Taylor, Peel and Sarno (2001) and Lothian and Taylor (2004). The value-added offered by this study is the conclusion that this evidence of non-linear adjustment remains even once explicit account has been taken of the impact of supply shocks to the real exchange rate due to productivity and Terms of Trade innovations. But it also appears fair to conclude that overall evidence of remaining residual non-linearity is weaker than reported than some other studies, perhaps reflecting our explicit capture of productivity and Terms of Trade effects on the path of the real exchange rate.

In conclusion, though, there does appear to be sufficient evidence from the range of diagnostic test results reported above of residual non-linearity within optimal linear VECMs to justify estimation of a set of ESTAR models, and it is to this analysis that we now turn.

2.5.2 Smooth Transition Autoregression Analysis

Our prior assumption is that the speed of adjustment of the real exchange rate around its equilibrium value is dependent upon only the magnitude, and not the direction of any divergence. This assumption is analysed using the Granger-Teräsvirta (1996) test outlined above (equations (22) to (24)) that aims to differentiate between ESTAR and LSTAR non-linearity within the residuals from estimated VECMs. These tests are based upon a Taylor expansion of underlying STAR models; as such they represent an approximate guide to model selection, rather than a definitive indicator. Test results are reported in Table 8, and suggest that our prior assumption is correct for yen-dollar and mark-sterling; these exchange rates should be modelled as ESTAR processes. For mark-dollar and yen-sterling results suggest an LSTAR model that incorporates asymmetric adjustment around Fair Value. This seems unintuitive, as discussed above, but is consistent with the results of Enders and Dibooglu (2001) and Leon and Najarian (2003). However, estimation results-not reported-indicate that an LSTAR specification is not appropriate for either exchange rate; estimated smoothness parameters are wrongly signed and insignificant. Consequently, we will maintain the assumption that all four exchange rates are ESTAR processes and test the validity and explanatory power of estimated models subsequently. We also maintain the hypothesis that neglected non-linearity exists in the adjustment of the exchange rate around equilibrium alone, and not in the estimated equilibrium series itself. Consequently, our chosen transition variable is the equilibrium correction term from the linear VECMs above. This approach represents an accurate interpretation of the intuition underlying the STAR methodology, and is supported by the evidence of Escribano and Miri (1996) and Balke and Fomby (1997), as discussed above.

Table 9 presents the results of our non-linear modelling. Estimation of ESTAR equations is undertaken by non-linear least squares, resulting in consistent and asymptotically normal estimators. Consistent with Granger and Teräsvirta (1993), we adopt a starting value for each ESTAR equation of one for the transition parameter in each of these three specifications, and scale this parameter by the standard deviation of the transition variable, ECM_t. This approach provides standardised results that are comparable across each of the four exchange rates that we examine and also allows the models to solve more easily. Results are not materially different using alternative starting values.

Estimated transition parameters are significantly different from zero for all four exchange rates, on the basis of reported t-statistics. Although t-statistics should be treated with a degree of caution, due to the presence of a unit root in each of our dependent and explanatory series, we verify these results using a bootstrap simulation incorporating 10,000 interations; these results (reported in square parentheses in Table 9) also indicate that the estimated transition parameters are significantly different from zero at a 1% significance level. The transition between the two states specified by the ESTAR model-unit root and mean reversion-occurs most rapidly for mark-dollar (two quarters) and yen-sterling (three quarters). However, transition speeds are relatively fast for the other two exchange rates as well-with both below two years-consistent with the findings of Lothian and Taylor (2004). Consequently, for all these exchange rates our estimation results suggest that market participants reach a consensus view concerning the extent of misalignments relatively quickly-compared to half-live estimates of traditional PPP models-and exploit consequent arbitrage opportunities causing the real exchange rate to converge back towards its equilibrium level.

It is important to test for the presence of additional neglected nonlinearity in ESTAR models, as well as serial independence and parameter constancy. Failure to pass any of these three diagnostic tests is taken as indicative of misspecification and should lead to model respecification. We examine for neglected non-linearity using a form of the Ramsey Reset test, whereby the residuals from ESTAR models are regressed on higher powers of the fitted values of the optimal ESTAR model in addition to the original independent regressors. The joint null hypothesis of this test is that the coefficient of each of these fitted terms is zero. As Table 10 highlights, there is no evidence of neglected non-linearity for any of the four exchange rates in our study. Similarly, for Table 11 there appears to be no significant evidence of residual serial correlation or parameter instability at traditional significance levels. Overall, therefore, the ESTAR models appear to be well specified.

The final diagnostic tests that we consider assess whether the selected ESTAR models encompass their linear equivalents. This assessment is important as it allows a direct assessment of the extent to which adopting a nonlinear augmented PPP specification is actually able to improve upon the explanatory power of the linear alternative; in a sense, this test assesses the economic significance of the statistical improvement that we have reported so far in moving from a linear to a non-linear specification. We perform this assessment in two ways. First, and following the approach of Skalin and Teräsvirta (1998), we assess the accuracy of the linear and non-linear specification by calculating the ratio S_{NL}/S_{LIN}, where S_{NL} is the estimated standard deviation of the residuals from the preferred ESTAR model and S_{LIN} is equal to

$$S_{LIN} = \sqrt{\frac{1}{T-k}\sum \hat{\varepsilon}_t^2} \tag{34}$$

where $\hat{\varepsilon}_t$ are the residuals from the optimal linear VECM. Second, by introducing into the ESTAR model for each exchange rate the linear equilibrium correction term and assessing the statistical significance of both the non-linear smoothness parameter and the linear equilibrium correction parameter in this augmented regression; a statistically significant smoothness parameter and an insignificant equilibrium correction parameter would imply that the ESTAR model encompasses the linear alternative, and vice versa. From Table 12, there appears to be little evidence that ESTAR models actually do encompass their linear alternatives: using the first test, only for yen-sterling do we see a reduction in the residual standard deviation of approximately 5%; and using the second test only the results for mark-sterling suggest the ESTAR model encompasses the linear alternative. For none of our exchange rate do the results of both tests unequivocally indicate that the non-linear ESTAR model encompasses the linear alternative. Given evidence from other studies that important nonlinearities exist within real exchange rates, this result appears disappointing. There are a number of potential explanations. First, this result may reflect our use of quarterly data (Taylor, 2000). We opted to use this frequency to ensure that the path of estimated equilibria was relatively smooth, in turn minimising the impact of changes in equilibria on the turnover, transaction costs and performance of investment portfolios. However, as real exchange rate data are sampled on a daily frequency this choice may have eliminated much of the inherent non-linear structure in these series. Moving to a monthly frequency may therefore improve the performance of our ESTAR models, albeit at the cost of some deterioration in the potential performance of associated investment portfolios. A second possibility reflects our augmentation of the traditional PPP relationship with supply related variables; the payoff from augmenting PPP with real variables is a faster convergence of exchange rates toward estimated equilibria in the wake of unanticipated shocks. But this payoff may have come at the cost, again, of undermining the inherent non-linear structure within exchange rate series. Third, deviations from equilibrium during our sample period have typically been very persistent; extending our data span may allow the inherent non-linear structure in exchange rates series to emerge more clearly in estimated ESTAR models. Fourth, it may be that our choice of an ESTAR framework is inappropriate and that exchange rates exhibit alternative non-linear forms, when account is taken of the impact of supply related variables. We leave these issues to future research, but suggest that addressing them may facilitate a general rehabilitation of ESTAR models, consistent with the results of other research in this area (Michael, Nobay, and Peel, 1997; O'Connell and Wei, 1997; Taylor and Peel, 2000; Taylor, Peel and Sarno, 2001; Kilian and Taylor, 2002; Sarno, Taylor and Chowdhury, 2002; Leon and Najarian, 2003).

2.6 Conclusion

The persistence of PPP deviations has long been an important focus of academic research. Traditional estimates of the half-life of shocks to PPP lie in the range three to five years. Efforts to explain this persistence have concentrated upon a number of areas, including: the presence of statistical biases in half-live estimates (Taylor, 2000; Imbs, Mumtaz, Ravn and Rey, 2002; Cashin and McDermott, 2003; Chen and Engel, 2004); imperfections in market structure (Krugman, 1986; Knetter, 1993); supply and demand shocks, including due to the HBS effect and shifts in the Terms of Trade (Asea and Cordon, 1994; Froot and Rogoff, 1994; Sarno and Taylor, 2002); and non-linear adjustment dynamics around linear (or log-linear) equilibria (Michael, Nobay, and Peel, 1997; O'Connell and Wei, 1997; Taylor and Peel, 2000; Taylor, Peel and Sarno, 2001; Kilian and Taylor, 2002; Sarno, Taylor and Chowdhury, 2002; Leon and Najarian, 2003). These various research strands have in turn provided a wide range of results-partly depending upon the length of data span and exchange rates analysed, estimation methodologies adopted, and so on-without any being entirely conclusive, indicating that the question of persistence is complex. Our approach in this study has been to examine the combined impact upon persistence of two of these factors during the floating rate era: supply shocks and non-linear adjustment dynamics. We are unaware of any other research that makes a similar contribution to the literature using this sample period; although see Lothian and Taylor (2004) for a complementary long-span study.

Consistent with the existing literature, our findings provide mixed reading. First, although productivity and Terms of Trade shocks do appear important to the determination of exchange rates during our sample period, we find only limited evidence in favour of the HBS effect. Second, the magnitude of equilibrium correction parameters within estimated linear VECMs suggest that augmentation of traditional PPP models with these supply side variables can greatly reduce the persistence of real exchange rate deviations from PPP. Third, evidence of residual non-linearity in these linear VECMs encourages us to model real exchange rate dynamics, incorporating the impact of supply side variables, as an ESTAR process. In turn, these models suggest that transition between unit root and mean reversion states embedded within the ESTAR framework occurs relatively quickly-within two years at the maximum-again suggesting that our approach is capable of substantially reducing the half-lives of PPP deviations. Fourth, although estimated ESTAR models are statistically well specified and significant, encompassing tests indicate little economic benefit is gained from replacing traditional linear VECMs incorporating supply variables with non-linear models of this form. We discuss a number of reasons that may help explain this last result, including our choice of a quarterly data frequency, augmentation of PPP with supply related variables and the persistence of deviations from equilibrium during our data sample. Overall, therefore, although we appear to have made some incremental progress towards explaining the persistence of PPP deviations in this study, we have also highlighted some possible issues with the recent thrust of fundamental-based research in this problem. The search goes on.

References

[1] Abauf, N. and P. Jorion (1990), Journal of Finance, Vol. 45, 1, 157-174.

[2] Alquist, R. & M.D. Chinn (2000), Productivity and the Euro-Dollar Exchange Rate Puzzle. NBER Working Paper # 8824.

[3] Asea, P.K. & W.M. Corden (1994), The Balassa-Samuelson Model: An Overview. Review of International Economics, vol. 2 , 191-200.

[4] Bailey, A., S. Millard & S. Wells (2001), Capital Flows and Exchange Rates. Bank of England Quarterly Bulletin. Autumn.

[5] Balassa, B. (1964), The Purchasing Power Parity Doctrine: A Reappraisal." Journal of Political Economy 72, 584-96.

[6] Balke, N. S. and T. B. Fomby (1997), Threshold Cointegration. International Economic Review, Vol. 38, 3, 627-645.

[7] Beveridge, S. and C.R. Nelson (1991), A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the "Business Cycle". Journal of Monetary Economics, 7, pp.151-74.

[8] Brock, W., D. Dechert, J. Sheinkman and B. LeBaron (1996). "A Test for Independence Based on the Correlation Dimension," Econometric Reviews, August 15(3), 197-235.

[9] Canzoneri, M.B., R.E. Cumby and B. Diba (1996), Relative Labor Productivity and the Real Exchange Rate in the Long Run: Evidence for a Panel of OECD Countries. National Bureau of Economic Research Working Paper # 5676.

[10] Canzoneri, M., R. Cumby, B. Diba and G. Eudey (1998), Trends in European Productivity: Implications for Real Exchange Rates, Real Interest Rates and Inflation Differentials. Oesterreichische Nationalbank Working Paper # 27.

[11] Cashin, P. and C.J. McDermott (2003), An Unbiased Appraisal of Purchasing Power Parity. IMF Staff Papers, Vol. 50, 3, 321-351.

[12] Cassel, G. (1918), Abnormal Deviations In International Exchanges. Economic Journal, 28, 413-415.

[13] Chen, L.L. and J. Devereux (2003), What Can US City Data Tell Us About Purchasing Power Parity? Journal of International Money and Finance, 22, 213-222.

[14] Chen, Y. and K. Rogoff (2002), Commodity Currencies and Empirical Exchange Rate Puzzles. International Monetary Fund Working Paper # WP/02/27.

[15] Chen, S.-S. and C. Engel (2004), Does Aggregation Bias Explain the PPP Puzzle? National Bureau of Economic Research Working Paper # 10304.

[16] Cheung, Y.-W. and K. S. Lai (1993), Long-Run Purchasing Power Parity during the Recent Float. Journal of International Economics, 34, 181-192.

[17] Cheung, Y.-W. , M.D. Chinn, M.D. and E. Fujii (1999), Market Structure and the Persistence of Sectoral Real Exchange Rates. International Journal of Finance and Economics 6, 2, 95-114.

[18] Cheung, Y.-W. and K. S. Lai (2000), On Cross-Country Differences in the Persistence of Real Exchange Rates. Journal of International Economics, 50, 375-397.

[19] Chinn, M.D. and L. Johnston (1997), Real Exchange Rate Levels, Productivity and Demand Shocks: Evidence from a Panel of Fourteen Countries. IMF Working Paper #WP/97/66.

[20] Chinn, M.D. (1998), The Usual Suspects: Productivity and Demand Shocks and Asia-Pacific Real Exchange Rates. Oesterreichische Nationalbank Working Paper # 31.

[21] Chinn, M.D. (2000), The Empirical Determinants of the Euro: Short and Long Run Perspectives. Paper Prepared For the Deutsche Bundesbank Conference, Equilibrium Exchange Rates of the Euro.

[22] Clarida, R.H. and M.P. Taylor (1997), The Term Structure of Forward Exchange Rate Premiums and the Forecastability of Spot Exchange Rates: Correcting the Errors. Review of Economics and Statistics, 79, 353-370.

[23] Closterman, J. and B. Schnatz (2000), The Determinants of the Euro-Dollar Exchange Rate: Synthetic Fundamentals and a Non-Existing Currency. Bundesbank Discussion Paper 2/00.

[24] Coakley, J., R.P. Flood & M.P. Taylor (2002), Long-Run Relative Purchasing Power Parity: The First Tests. University of Warwick mimeo.

[25] Crucini, M.J., C.I. Telmer and M. Zachariadis (2000), Cross-Sectional Variation in European Real Exchange Rates: 1975-1990. Vanderbilt University Mimeo.

[26] Crucini, M.J. and M. Shintani (2001), Persistence in Law-of-One-Price Deviations: Evidence from Micro-Data. Vanderbilt University mimeo.

[27] Daníelsson, J., R. Payne and J. Luo (2002), Exchange Rate Determination and Inter-Market Order Flow Effects. LSE Financial Markets Group Mimeo.

[28] DeGregorio, J. and H. Wolf (1994), Terms of Trade, Productivity and the Real Exchange Rate. National Bureau of Economic Research Working Paper # 4807.

[29] DeLoach, S. (2001), More Evidence in Favor of the Balassa-Samuelson Hypothesis. Review of International Economics, 9(2), 336-342.

[30] Dornbusch, R. (1988), Exchange Rates and Inflation. The MIT Press.

[31] Eitrheim, Ø & T. Teräsvirta (1996), Testing the Adequacy of Smooth Transition Autoregressive Models. Journal of Econometrics, 74, 59-75.

[32] Enders, W. (2003), Applied Econometric Time Series. Wiley.

[33] Enders, W. and S. Dibooglu (2001), Long Run Purchasing Power Parity With Asymmetric Adjustment. Southern Economic Journal Vol. 68, 2, 433-445.

[34] Engel, C. (1993), Real Exchange Rates and Relative Prices: An Empirical Investigation. Journal of Monetary Economics, 32, 35-50.

[35] Escribano, A. and S. Miri (1996), Nonlinear Cointegration and Nonlinear Error Correction Models. Working Papers Series in Statistics and Econometrics, 96-54. Universidad Carlos III de Madrid.

[36] EuroMoney (2002), 2nd International Forex Forum. Euromoney.

[37] Evans, M. and R.K. Lyons (2002), Order Flow and Exchange Rate Dynamics. Journal of Political Economy, February, 170-180.

[38] Faruqee, H. (1995), Long-Run Determinants of the Real Exchange Rate: A Stock-low Perspective. International Monetary Fund Staff Papers, Vol. 42, 80-107.

[39] Franses, P.H. and D. van Dijk (2000), Non-Linear Time Series Models in Empirical Finance. Cambridge University Press.

[40] Froot, K.A. and K. Rogoff (1994), Perspectives on PPP and Long-Run Real Exchange Rates. National Bureau of Economic Research Working Paper Working Paper #4952.

[41] Granger, C.W.J. (1993), Co-integrated Variables and Error Correcting Models. UCSD Discussion Paper.

[42] Granger, C.W.J. and T. Teräsvirta (1996), Modelling Non-Linear Economic Relationships. Oxford University Press.

[43] Harrod, R. (1933), International Economics. James Nisbet and Cambridge University Press.

[44] Hendry, D., and N. R. Ericcson (1991), An Econometric Analysis of UK Money Demand in Monetary Trends in the United States and the United Kingdom by Milton Friedman and Anna J. Schwartz. American Economic Review, 81, 1, 8-38.

[45] Imbs, J., H. Mumtaz, M. O. Ravn and H. Rey (2002), PPP Strikes Back: Aggregation and the Real Exchange Rate. National Bureau of Economic Research Working Paper Working Paper #9372.

[46] International Monetary Fund (2001), The Information Technology Revolution. World Economic Outlook. October.

[47] International Monetary Fund (2002), The Rise of the Sterling Exchange Rate. In United Kingdom: Selected Issues. IMF Country Report # 02/46. March.

[48] Ito, T., P. Isard and S. Symansky (1997), Economic Growth and Real Exchange Rates: An Overview of the Balassa-Samuelson Hypothesis in Asia. National Bureau of Economic Research Working Paper # 5979.

[49] Isard, P. (1977), How Far Can We Push the Law of One Price. American Economic Review. Vol. 67, 942-948.

[50] Johansen, S. (1988), Statistical Analysis of Cointegrating Vectors. Journal of Economic Dynamics and Control 12, 231-54.

[51] Kilian, L. and M. P. Taylor (2002), Why is it so Difficult to Beat the Random Walk Forecast of Exchange Rates? Journal of International Economics, 60, 85-107.

[52] Knetter, M.M. (1993), International Comparisons of Pricing-to-Market Behavior. American Economic Review, 83, 3, 473-486.

[53] Koen, V., L. Boone, A. de Serres and N. Fuchs (2001), Tracking the Euro. OECD Economics Department Working Paper # 298.

[54] Krugman, P. (1986), Pricing to Market when the Exchange Rate Changes. NBER Working Paper #1926.

[55] Leon, H. & S. Najarian (2003), Asymmetric Adjustment and Nonlinear Dynamics in Real Exchange Rates. IMF Working Paper WP/03/159. www.imf.org.

[56] Liu, Y. (2001), Modelling Mortgage Rate Changes with a Smooth Transition Error-Correction Model. Bank of Canada Working Paper # 2001-23.

[57] Lothian and M. P. Taylor (1996), Real Exchange Rate Behaviour: The Recent Float From the Behaviour of the Past Two Centuries. Journal of Political Economy, vol. 104, 3, 488-509.

[58] Lothian, J.R. and M.P. Taylor (2004) Real Exchange Rates over the Last Two Centuries: How Important is the Harrod-Balassa-Samuelson Effect? Unpublished University of Warwick mimeo.

[59] Lucas, R.E. (1982), Interest Rates and Currency Prices in a Two-Country World Journal of Monetary Economics, 10, 335-359.

[60] Meese, R.A. and K. Rogoff (1983a), Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample? Journal of International Economics, 14, 3-24

[61] Meese, R.A. and K. Rogoff (1983b), The Out-of-Sample Failure of Empirical Exchange Rate Models: Sampling Error or Misspecification? in Frenkel, J.A. (ed.), Exchange Rate and International Economics, University of Chicago Press

[62] Michael, P., R. Nobay, and D. A. Peel (1997), Transactions Costs and Nonlinear Adjustment in Real Exchange Rates: An Empirical Investigation. The Journal of Political Economy, Vol. 105, 4, 862-879.

[63] Mussa, M. (1984), The Theory of Exchange Rate Determination. In Exchange Rate Theory and Practice. NBER Conference Report, J.F.O. Bilson and R.C. Marston (eds), Chicago University Press.

[64] O'Connell, P.G.J. and S.-J. Wei (1997), The Bigger They Are , the Harder They Fall: How Price Differences Across US Cities Are Arbitraged, NBER Working Paper # 6089.

[65] O'Connell, P.G.J. (1998), The Overvaluation of Purchasing Power Parity. Journal of International Economics, 44, 1-19.

[66] Osbat, C., R. Rüffer and B. Schnatz (2003), The Rise of the Yen vis-à-vis the ("Synthetic") Euro: Is it Supported by Economic Fundamentals? European Central Bank Working Paper #224.

[67] Ramsey, J.B. (1969), Tests for Specification Errors in classical Linear Least Squares Regression Analysis. Journal of the Royal Statistical Society , Series B, 31, 350-371.

[68] Richardson, J.D. (1978), Some Evidence on Commodity Arbitrage and the Law of One Price. Journal of International Economics, 38, 339-360.

[69] Rogoff, K. (1996), The Purchasing Power Parity Puzzle, Journal of Economic Literature, Vol. 34, 647-668.

[70] Rogoff, K., K.A. Froot and M. Kim (2001), The Law of One Price Over 700 Years. International Monetary Fund Working Paper #WP/01/174.

[71] Sager, M.J., N. Nuttall and M. P. Taylor (2002), The Outlook for the US Dollar. European Pensions and Investment News, 100, i-iv. March.

[72] Sager, M.J. and M.P. Taylor (2004), The Role of Order Flow in Exchange Rate Forecasting. Unpublished University of Warwick Mimeo.

[73] Samuelson, P. (1964), Theoretical Notes on Trade Problems. Review of Economics and Statistics 46, 145-54.

[74] Sarno, L. and M.P. Taylor (2001), The Microstructure of the Foreign-Exchange Market: A Selective Survey of the Literature. Princeton Studies in International Economics, 89.

[75] Sarno, L. and M.P. Taylor (2002), The Economics of Exchange Rates. Cambridge: Cambridge University Press.

[76] Sarno, L., M. P. Taylor and I. Chowdhury (2002), Non-Linear Dynamics in Deviations from the Law of One Price: A Broad-Based Empirical Study. Centre for Economic Policy Research Discussion Paper # 3377.

[77] Schnatz, B., F. Vijselaar and C. Osbat (2003), Productivity and the ("Synthetic") Euro-Dollar Exchange Rate. European Central Bank Working Paper #225.

[78] Stockman, A.C. (1980), A Theory of Exchange Rate Determination. Journal of Political Economy, 88, 4, 673-698.

[79] Taylor, A.M. (2000), Potential Pitfalls for the Purchasing Power Parity Puzzle? Sampling and Specification Biases in Mean Reversion Tests of the Law of One Price. National Bureau of Economic Research Working Paper # 7577.

[80] Taylor, A.M. (2002), A Century of Purchasing Power Parity. The Review of Economics and Statistics, Vol. 84, 1, pp.139-50.

[81] Taylor, M.P. and D.A. Peel (2000), Non-Linear Adjustment, Long-Run Equilibrium and Exchange Rate Fundamentals. Journal of International Money and Finance, 19, 33-53

[82] Taylor, M.P., D. Peel and L. Sarno (2001), Non-Linear Mean Reversion in Real Exchange Rates: Towards a Solution to the Purchasing Power Parity Puzzles. International Economic Review, 42, 1015-42.

[83] Teräsvirta, T. and H.M. Anderson (1992), Characterising Nonlinearities in Business Cycles Using Smooth Transition Autoregressive Models. Journal of Applied Econometrics, Vol. 7, 119-136.

[84] Teräsvirta (1994), Specification, Estimation and Evaluation of Smooth Transition Autoregressive Model. Journal of the American Statistical Association, 89, 208:218.

[85] Teräsvirta (1998), Modelling Economic Relationships With Smooth Transition Regressions. In D.E.A Giles and A. Ullah (eds.), Handbook of Applied Economic Statistics. New York: Marcel Dekker, 507-552.

[86] Tong, H. (1990), Nonlinear Time Series: A Dynamic System Approach. Clarendon Press.

[87] Van Dijk, D. and P. H. Franses (2000), Non-Linear Error Correction Models for Interest Rates in the Netherlands. In Non-Linear Econometric Modelling in Time Series Analysis, eds. W. A. Barnett, D. F. Hendry, S. Hylleberg, T. Teräsvirta, D. Tjøstheim and A. Würtz. Cambridge: Cambridge University Press.

[88] Van Dijk, P. H. Franses and T. Teräsvirta, (2000), Smooth Transition Autoregressive Models – A Survey of Recent Developments. Econometric Institute Research Report EI2000-23/A.

Table 1: Augmented Dickey Fuller (ADF) Unit Root Tests

	$I(0)$			
Mark – Dollar	*Constant*	*Lag Length*	*Constant and Trend*	*Lag Length*
RER	−2.4785	4	−2.5414	4
BD_Man	0.0329	0	−1.3955	0
US_Man	1.2867	1	−1.0538	1
Man_Prod	2.5201	0	−2.4976	0
BD_Man_Whole	−0.6593	0	−2.7639	0
US_Man_Whole	1.6234	5	−1.5844	5
Man_Whole	−2.5622	0	−3.6747*	0
BD_ToT	−1.2604	1	−2.6753	3
US_ToT	−3.3632*	3	−3.5381*	3
ToT	−1.6472	1	−2.7585	1
BD_Gov	−2.2165	10	−3.0649	10
US_Gov	−2.4547	7	−2.5748	7
Gov_Bal	−2.6305	7	−3.6910*	9
Oil	−1.4364	5	−2.8457	3

Notes: Lag length selected to minimise AIC. * (**) indicates significance at 5% (1%) level. RER is the mark-dollar real exchange rate, defined in terms of CPIs. BD_Man and US_Man are manufacturing output per man in Germany-our proxy for the Euro Area-and the US. Man-Prod is the differential of these two series. BD_Man_Whole and US_Man_Whole are the ratio of manufacturing output per man to whole economy output per man in Germany and the US. BD_Tot and US_Tot are the Terms of Trade in Germany and the US, defined as the ratio of export to import prices for both countries. ToT is the relative Terms of Trade between these two countries. BD_Gov and US_Gov are government consumption as a percentage of nominal GDP in Germany and the US. Oil is the nominal price of oil, deflated by the US CPI. All series are in natural logarithms.

Table 1 (cont.): Augmented Dickey Fuller (ADF) Unit Root Tests

	$I(1)$			
	Constant	*Lag Length*	*Constant and Trend*	*Lag Length*
Mark – Dollar				
RER	−4.9346**	2	−4.9112**	2
BD_Man	−9.2442**	0	−9.2083**	0
US_Man	−7.8182**	0	−7.9997**	0
Man_Prod	−2.7281*	10	−2.7615	10
BD_Man_Whole	−10.6282**	0	−10.5876**	0
US_Man_Whole	−3.9210**	4	−3.1891*	12
Man_Whole	−11.2003**	0	−11.3526**	0
BD_ToT	−4.3593**	3	−4.3649**	3
US_ToT	−5.8906**	4	−5.9667**	4
ToT	−7.3449**	0	−7.3149**	0
BD_Gov	−5.5533**	9	−5.5199**	9
US_Gov	−9.2868**	4	−9.2620**	4
Gov_Bal	−4.1547**	9	−4.1211**	9
Oil	−5.6957**	4	−5.6613**	4

Notes: Lag length selected to minimise AIC. * (**) indicates significance at 5% (1%) level.

Table 1 (cont.): Augmented Dickey Fuller (ADF) Unit Root Tests

	$I(0)$			
	Constant	*Lag Length*	*Constant and Trend*	*Lag Length*
$Yen - Dollar$				
RER	-2.3516	3	-2.4621	3
JP_Man	-2.5077	9	-2.0553	9
Man_Prod	0.5533	1	-1.8331	1
JP_Man_Whole	-0.4255	12	-3.3852	12
Man_Whole	1.3881	10	-2.2511	5
JP_ToT	-2.6481	1	-2.7809	1
ToT	-2.3595	1	-2.4527	1

Notes: Lag length selected to minimise AIC. * (**) indicates significance at 5% (1%) level. RER is the yen-dollar real exchange rate, defined in terms of CPIs. JP_Man is manufacturing output per man in Japan. Man-Prod is the differential of JP_Man and US_Man. JP_Man_Whole is the ratio of manufacturing output per man to whole economy output per man in Japan. Man_Whole is the differential between JP_Man_Whole and US_Man_Whole. JP_Tot is the Terms of Trade in Japan, defined as the ratio of export to import prices. ToT is the relative Terms of Trade between Japan and the US. All series are in natural logarithms.

Table 1 (cont.): Augmented Dickey Fuller (ADF) Unit Root Tests

	$I(1)$			
	Constant	*Lag Length*	*Constant and Trend*	*Lag Length*
Yen – Dollar				
RER	-5.0254^{**}	2	-5.0611^{**}	2
JP_Man	-2.5470^{*}	12	-2.6688^{*}	12
Man_Prod	-2.1114^{*}	12	-2.7794^{*}	12
JP_Man_Whole	-2.8513^{*}	10	-3.5284^{*}	9
Man_Whole	-4.4859^{**}	4	-4.6434^{**}	4
JP_ToT	-6.0569^{**}	0	-6.0377^{**}	0
ToT	-6.6114^{**}	0	-6.5834^{**}	0

Notes: Lag length selected to minimise AIC. * (**) indicates significance at 5% (1%) level.

Table 1 (cont.): Augmented Dickey Fuller (ADF) Unit Root Tests

	$I(0)$			
	Constant	*Lag Length*	*Constant and Trend*	*Lag Length*
Mark – Sterling				
RER	−2.5755	0	−2.4537	0
BD_Man	0.7191	0	−2.0793	0
UK_Man	−1.0926	0	−1.0073	0
Man_Prod	−0.6917	0	−1.1248	0
BD_Man_Whole	−0.1643	8	−1.6975	8
UK_Man_Whole	−1.1653	2	−1.5392	2
Man_Whole	−2.3593	8	−2.1526	8
BD_ToT	−1.1692	1	−2.6080	1
UK_ToT	−1.2328	12	−1.4840	12
ToT	−1.8387	9	−1.4085	9

Notes: Lag length selected to minimise AIC. * (**) indicates significance at 5% (1%) level. RER is the mark-sterling real exchange rate, defined in terms of CPIs. BD_Man and UK_Man are manufacturing output per man in Germany-our proxy for the Euro Area-and the UK. Man-Prod is the differential of these series. BD_Man_Whole and UK_Man_Whole are the ratio of manufacturing output per man to whole economy output per man in Germany and the UK. Man_Whole is the differential of these two series. BD_Tot and UK_Tot are the Terms of Trade in Germany and the UK, defined as the ratio of export to import prices for both countries. ToT is the relative Terms of Trade between these two countries. All series are in natural logarithms.

Table 1 (cont.): Augmented Dickey Fuller (ADF) Unit Root Tests

	$I(1)$			
	Constant	*Lag Length*	*Constant and Trend*	*Lag Length*
Mark – Sterling				
RER	−8.8477**	0	−8.8471**	0
BD_Man	−8.7435**	0	−8.8080**	0
UK_Man	−5.5662**	1	−9.1617**	0
Man_Prod	−8.9925**	0	−9.2366**	0
BD_Man_Whole	−2.6646*	7	−4.9454**	4
UK_Man_Whole	−4.9245**	1	−4.9419**	1
Man_Whole	−2.0530*	7	−2.3997*	7
BD_ToT	−4.0395**	3	−4.0077**	3
UK_ToT	−2.7252*	11	−3.2743*	11
ToT	−3.3710*	8	−3.4804*	8

Notes: Lag length selected to minimise AIC. * (**) indicates significance at 5% (1%) level.

Table 1 (cont.): Augmented Dickey Fuller (ADF) Unit Root Tests

	$I(0)$			
	Constant	*Lag Length*	*Constant and Trend*	*Lag Length*
Yen − Sterling				
RER	−2.1658	1	−2.4601	1
JP_Man	−0.8649	9	−2.0712	9
UK_Man	−0.5165	0	−1.1980	0
Man_Prod	−1.5808	2	−2.9433	2
JP_Man_Whole	−2.5028	12	−2.8677	12
UK_Man_Whole	−1.1653	2	−1.5392	2
Man_Whole	−1.7278	12	−2.014	12
JP_ToT	−2.4344	9	−2.4363	9
UK_ToT	−1.3044	12	−1.5326	12
ToT	−2.5112	9	−2.1620	9

Notes: Lag length selected to minimise AIC. * (**) indicates significance at 5% (1%) level. RER is the yen-sterling real exchange rate, defined in terms of CPIs. JP_Man and UK_Man are manufacturing output per man in Japan and the UK. Man-Prod is the differential of these two series. JP_Man_Whole and UK_Man_Whole are the ratio of manufacturing output per man to whole economy output per man in Japan and the UK. Man_Whole is the differential of these two series. JP_Tot and UK_Tot are the Terms of Trade in Japan and the UK, defined as the ratio of export to import prices for both countries. ToT is the relative Terms of Trade between these two countries. All series are in natural logarithms.

Table 1 (cont.): Augmented Dickey Fuller (ADF) Unit Root Tests

	$I(1)$			
	Constant	*Lag Length*	*Constant and Trend*	*Lag Length*
Yen – Sterling				
RER	-7.7380^{**}	0	-7.6920^{**}	0
JP_Man	-3.5543^{*}	8	-3.5110^{*}	8
UK_Man	-5.5662^{**}	1	-8.9672^{**}	0
Man_Prod	-4.3826^{**}	1	-4.3461^{**}	1
JP_Man_Whole	-2.3473	11	-2.3257	11
UK_Man_Whole	-4.9245^{**}	1	-4.9419^{**}	1
Man_Whole	-1.8312	11	-2.0209	11
JP_ToT	-2.8959^{*}	8	-2.8695^{*}	8
UK_ToT	-3.4945^{**}	11	-3.9705^{**}	11
ToT	-2.8703^{*}	8	-2.8765^{*}	8

Notes: Lag length selected to minimise AIC. * (**) indicates significance at 5% (1%) level.

Table 2: Johansen Cointegration Analysis

Mark – Dollar

RER, BD_Man_1, US_Man, BD_ToT
No Deterministic Trend; Lag Length: 4

No. Hypothesised CEs	*Eigenvalue*	*Trace Statistic*	5% *Critical Value*	1% *Critical Value*
*None***	0.2326	67.1769	53.12	60.16
*At Most One**	0.1891	38.8407	34.91	41.07
At Most Two	0.1227	16.3997	19.96	24.60
At Most Three	0.0220	2.3858	9.24	12.97

Yen – Dollar

RER, Man_whole, JP_ToT
Linear Deterministic Trend; Lag Length: 3

No. Hypothesised CEs	*Eigenvalue*	*Trace Statistic*	5% *Critical Value*	1% *Critical Value*
*None***	0.3014	49.9612	42.44	48.45
At Most One	0.1055	18.0321	25.32	30.45
At Most Two	0.0870	8.1024	12.25	16.26

Notes: * (**) indicates significance at 5% (1%) level.

Table 2 (cont.): Johansen Cointegration Analysis

Mark – Sterling

RER, Man_prod, ToT
Linear Deterministic Trend; Lag Length: 5

No. Hypothesised CEs	*Eigenvalue*	*Trace Statistic*	5% *Critical Value*	1% *Critical Value*
*None**	0.2205	31.7012	29.68	35.65
At Most One	0.0666	8.7816	15.41	20.04
At Most Two	0.0261	2.4360	3.76	6.65

Yen – Sterling

RER, UK_Man, JP_Man, UK_ToT
Linear Deterministic Trend; Lag Length: 4

No. Hypothesised CEs	*Eigenvalue*	*Trace Statistic*	5% *Critical Value*	1% *Critical Value*
*None***	0.2715	75.2674	62.99	70.05
*At Most One**	0.2145	45.4813	42.44	48.45
At Most Two	0.1724	22.7823	25.32	30.45
At Most Three	0.0517	4.9948	12.25	16.26

Notes: * (**) indicates significance at 5% (1%) level.

Table 3: Johansen Cointegration Coefficients

Mark – Dollar		*Yen – Dollar*	
BD_Man	−2.1001 (0.3955)	Man_Whole	2.0102 (0.4473)
US_Man	2.8111 (0.4073)		
BD_Tot	−2.4998 (0.2120)	JP_Tot	−2.1564 (0.2103)
		$Trend$	−0.0211 (0.0045)

Notes: Cointegrating parameter estimates derived from Johansen cointegration test. Standard errors in parentheses. Inclusion of a time trend is intended to capture supply side effects not captured by real variables included in our analysis.

Table 3 (Cont.): Johansen Cointegration Coefficients

Mark – Sterling		*Yen – Sterling*	
Man_Prod	0.6643 (0.2441)	JP_Man	1.1559 (0.2634)
		UK_Man	−1.9675 (0.3137)
Tot	0.4301 (0.2081)	UK_Tot	4.0991 (0.7078)
		$Trend$	0.0108 (0.0026)

Notes: Cointegrating parameter estimates derived from Johansen cointegration test. Standard errors in parentheses.

Table 4 : Linear VECM Analysis

Mark − Dollar		*Yen − Dollar*	
ECM	−0.1955 (0.0459)	ECM	−0.1380 (0.0319)
$\Delta RER(-2)$	−0.1987 (0.0779)	$\Delta RER(-2)$	−0.1882 (0.0875)
$\Delta RER(-4)$	0.1425 (0.0739)		
$\Delta US\_Man(-3)$	1.6969 (0.5550)		
$\Delta BD\_ToT(-3)$	−0.8618 (0.3053)		
$Adj.R^2$	0.2437	$Adj.R^2$	0.1190
SSR	0.2968	SSR	0.3461
AIC	−2.9660	AIC	−2.6665
JB^*	0.0088	JB^*	0.3141
$Breusch^*$	0.2289	$Breusch^*$	0.5626
$Arch(1)^*$	0.2095	$Arch(1)^*$	0.4701
$Arch(4)^*$	0.5258	$Arch(4)^*$	0.3618
$Ramsey$	1.5282	$Ramsey$	0.0581

Notes: The dependent variable is the quarterly change in the real exchange rate, Δq_t. ECM is the equilibrium correction mechanism derived from Johansen cointegration analysis, according to the Ganger Representation Theorem (Granger, 1983). Standard errors in parentheses. JB is the Jarque-Bera test for residual normality. Breusch is the Breusch-Godfrey test for residual serial autocorrelation. ARCH is the ARCH-LM test for residual heteroskedasticity, assuming one and four lags. Ramsey is the Ramsey Reset test, where we include quadratic and cubic terms of the fitted values of the VECM. * denotes p-value.

Table 4 (cont.): Linear VECM Analysis

Mark − Sterling		*Yen − Sterling*	
ECM	−0.1221	*ECM*	−0.1683
	(0.0318)		(0.0694)
		$\Delta RER(-1)$	0.2957
			(0.0922)
$\Delta Man\_prod(-4)$	−0.6798		
	(0.2690)		
$\Delta ToT(-3)$	−0.6568		
	(0.1981)		
$Adj.R^2$	0.1824	$Adj.R^2$	0.1044
SSR	0.1651	*SSR*	0.3485
AIC	−3.4423	*AIC*	−2.7603
JB^*	0.1911	JB^*	0.7890
$Breusch^*$	0.7854	$Breusch^*$	0.5905
$Arch(1)^*$	0.7556	$Arch(1)^*$	0.3119
$Arch(4)^*$	0.7549	$Arch(4)^*$	0.7317
Ramsey	0.8420	*Ramsey*	0.3752

Notes: The dependent variable is the quarterly change in the real exchange rate, Δq_t. ECM is the equilibrium correction mechanism derived from Johansen cointegration analysis, according to the Ganger Representation Theorem (Granger, 1983). Standard errors in parentheses. JB is the Jarque-Bera test for residual normality. Breusch is the Breusch-Godfrey test for residual serial autocorrelation. ARCH is the ARCH-LM test for residual heteroskedasticity, assuming one and four lags. Ramsey is the Ramsey Reset test, where we include quadratic and cubic terms of the fitted values of the VECM. * denotes p-value.

Table 5 : Autocorrelation Function from Optimal Linear VECMs

	$Mark - Dollar$	$Yen - Dollar$	$Mark - Sterling$	$Yen - Sterling$
ρ_1	0.218	0.455	0.748	0.298
ρ_2	0.420	0.415	0.791	0.558
ρ_3	0.388	0.606	0.577	0.670
ρ_4	0.552	0.338	0.740	0.797
ρ_5	0.685	0.259	0.702	0.855
ρ_6	0.689	0.215	0.711	0.885
ρ_7	0.763	0.300	0.804	0.825
ρ_8	0.840	0.300	0.833	0.889
ρ_9	0.651	0.098	0.682	0.934
ρ_{10}	0.520	0.027	0.670	0.931
ρ_{11}	0.602	0.040	0.482	0.936
ρ_{12}	0.578	0.060	0.353	0.954

Notes: Table reports p-values for autocorrelation function calculated using residuals from optimised VECMs.

Table 6 : BDS Tests

	Mark – Dollar			*Yen – Dollar*		
	Epsilon			*Epsilon*		
	0.25	0.50	0.75	0.25	0.50	0.75
Dimension						
2	0.6000	0.1506	0.0040	0.9030	0.9996	0.8800
3	0.5090	0.3192	0.0926	0.6592	0.8232	0.7820
4	0.4362	0.0442	0.0064	0.9594	0.3108	0.7972
5	0.7396	0.0426	0.0294	0.3016	0.2520	0.9374
6	0.3268	0.1920	0.0596	0.3224	0.1820	0.5336

Notes: The BDS test statistic tests for time based dependence in a series, including non-linearity, against the null hypothesis that a series is i.i.d., as described in the text. Epsilon $\{\varepsilon\}$ is the distance used for testing proximity of the data points and is calculated so as to ensure a certain fraction of the total number of pairs of points in the sample lie within $\{\varepsilon\}$ of each other. We run the test over three different values of epsilon to verify the robustness of test results. Dimension is the number of consecutive data points included in the test. As the BDS test may be different from the asymptotic normal distribution in small samples or in series that have unusual distributions we report bootstrapped p-values for the test statistic on the basis of 10,000 repetitions.

Table 6 (cont.) : BDS Tests

	Mark – Sterling			*Yen – Sterling*		
	Epsilon			*Epsilon*		
	0.25	0.50	0.75	0.25	0.50	0.75
Dimension						
2	0.2370	0.6690	0.6908	0.6712	0.2006	0.9900
3	0.1968	0.4444	0.8856	0.2666	0.1152	0.7000
4	0.2836	0.3558	0.7398	0.3338	0.1014	0.6492
5	0.5178	0.5100	0.6076	0.9786	0.3058	0.6482
6	0.6300	0.6578	0.4394	0.7296	0.4778	0.4360

Notes: The BDS test statistic tests for time based dependence in a series, including non-linearity, against the null hypothesis that a series is i.i.d., as described in the text. Epsilon $\{\varepsilon\}$ is the distance used for testing proximity of the data points and is calculated so as to ensure a certain fraction of the total number of pairs of points in the sample lie within $\{\varepsilon\}$ of each other. We run the test over three different values of epsilon to verify the robustness of test results. Dimension is the number of consecutive data points included in the test. As the BDS test may be different from the asymptotic normal distribution in small samples or in series that have unusual distributions we report bootstrapped p-values for the test statistic on the basis of 10,000 repetitions.

Table 7 : Granger Teräsvirta Linearity Tests

	Mark – Dollar	*Yen – Dollar*	*Mark – Sterling*	*Yen – Sterling*
1	0.2407	0.9049	0.9238	0.0816
2	0.1915	0.4695	0.7464	0.0031
3	0.3562	0.6093	0.3250	0.0029
4	0.2055	0.4711	0.2582	0.0707
5	0.4937	0.0661	0.1083	0.0406
6	0.3950	0.1218	0.0088	0.0595
7	0.2164	0.0417	0.0109	0.0317
8	0.2400	0.1120	0.0312	0.0265

Notes: Table reports p-values for Granger-Teräsvirta (1996) linearity tests on the residuals from the optimal VECM of each exchange rate.

Table 8 : ESTAR vs. LSTAR Tests

$H0$	$Mark - Dollar$	$Yen - Dollar$	$Mark - Sterling$	$Yen - Sterling$
3	0.2588	0.0191	0.0406	0.2447
2	0.1119	0.9247	0.3849	0.4576
1	0.7001	0.0085	0.0096	0.0028

Notes: Table calculates p-values for a series of nested F-tests based upon the auxiliary regression proposed by Granger and Teräsvirta (1993) and Teräsvirta (1994, 1998), as discussed in the text.

Table 9 : ESTAR Models

Mark – Dollar		*Yen – Dollar*	
γ $(d=2)$	-0.3482	γ $(d=7)$	-0.4394
	(0.0767)		(0.1445)
	[0.0000]		[0.0000]
$\Delta RER(-2)$	-0.2657	$\Delta RER(-2)$	-0.1308
	(0.0794)		(0.1050)
$\Delta RER(-4)$	0.1086		
	(0.0774)		
$\Delta US\_Man(-3)$	1.4891		
	(0.4405)		
$\Delta BD\_ToT(-3)$	-0.9644		
	(0.3383)		
$Adj.R^2$	0.2517	$Adj.R^2$	0.0977
SSR_{NL}	0.2899	SSR_{NL}	0.3273
SSR_{NL}/SSR_{LIN}	0.97	SSR_{NL}/SSR_{LIN}	0.94
AIC	-2.9693	AIC	-2.6495
JB^*	0.0437	JB^*	0.3318
$Breusch^*$	0.3929	$Breusch$	0.4645
$Arch(1)^*$	0.6372	$Arch(1)$	0.3259
$Arch(4)^*$	0.9182	$Arch(4)$	0.7251

Notes: Standard errors in parentheses. The optimal value of d, the delay parameter, is calculated over a range of values, $1 \leqslant d \leqslant D$. A significant rejection of the null hypothesis for any value of d is interpreted as indicative of the presence of a non-linear STAR process within the residuals of the estimated linear VECM. If two or more values of d are significant, the optimal value of d, $\hat{d}$, is chosen to be the one that minimises the p-value of the test statistic. Bootstrapped p-values of estimated transition coefficient reported in square brackets. SSR_{NL} is the sum of squared residuals from the ESTAR equations for each exchange rate; SSR_{LIN}.is the sum of squared residuals from the linear VECM. Other diagnostic tests as detailed above. * denotes p-value.

Table 9 (cont.) : ESTAR Models

Mark – Sterling		*Yen – Sterling*	
γ $(d = 6)$	-0.4027 (0.0719) [0.0000]	γ $(d = 3)$	-1.1981 (0.4665) [0.0022]
		$\Delta RER(-1)$	0.1829 (0.0717)
$\Delta Man\_prod(-4)$	-0.6302 (0.2748)		
$\Delta ToT(-3)$	-0.4125 (0.2699)		
$Adj.R^2$	0.1509	$Adj.R^2$	0.1083
SSR_{NL}	0.1476	SSR_{NL}	0.3008
SSR_{NL}/SSR_{LIN}	0.89	SSR_{NL}/SSR_{LIN}	0.86
AIC	-3.4716	AIC	-2.8416
Breusch	0.1035	*Breusch*	0.2124
$Arch(1)$	0.6724	$Arch(1)$	0.0874
$Arch(4)$	0.7061	$Arch(4)$	0.3942
Jarque-Bera	0.0518	*Jarque-Bera*	0.6144

Notes: Standard errors in parentheses. The optimal value of d, the delay parameter, is calculated over a range of values, $1 \leqslant d \leqslant D$. A significant rejection of the null hypothesis for any value of d is interpreted as indicative of the presence of a non-linear STAR process within the residuals of the estimated linear VECM. If two or more values of d are significant, the optimal value of d, $\hat{d}$, is chosen to be the one that minimises the p-value of the test statistic. Bootstrapped p-values of estimated transition coefficient reported in square brackets. SSR_{NL} is the sum of squared residuals from the ESTAR equations for each exchange rate; SSR_{LIN}.is the sum of squared residuals from the linear VECM. Other diagnostic tests as detailed above. * denotes p-value.

Table 10 : Eitrheim and Teräsvirta Tests for Residual Non-Linearity in ESTAR Models

	$Mark - Dollar$	$Yen - Dollar$	$Mark - Sterling$	$Yen - Sterling$
1	0.1264	0.8883	0.9602	0.8091
2	0.1263	0.2993	0.8899	0.4751
3	0.2384	0.5552	0.6640	0.1808
4	0.1496	0.6807	0.3753	0.7336
5	0.4541	0.1758	0.4996	0.3479
6	0.1552	0.3203	0.1468	0.5685
7	0.4250	0.6395	0.1202	0.4618
8	0.5138	0.8290	0.3142	0.2288

Notes: Table reports p-values derived from tests described in the text (equation (27)), as proposed by Eitrheim and Teräsvirta (1996).

Table 11 : Eitrheim and Teräsvirta Tests for Serial Independence and Parameter Constancy of ESTAR Models

	$Mark - Dollar$	$Yen - Dollar$	$Mark - Sterling$	$Yen - Sterling$
Serial Independence				
	0.9935	0.0793	0.1011	0.0654
Parameter Constancy				
	0.9999	0.9999	0.9999	0.9999

Notes: Table reports p-values derived from tests described in the text (equations (25), (26) (28) and (29)), as proposed by Eitrheim and Teräsvirta (1996).

Table 12 : Encompassing Tests

	$Mark - Dollar$	$Yen - Dollar$	$Mark - Sterling$	$Yen - Sterling$
S_L	0.0527	0.0630	0.0422	0.0574
S_{NL}	0.0539	0.0630	0.0426	0.0602
S_L/S_{NL}	0.9778	0.9991	0.9928	0.9535
γ	0.2212	0.2451	0.0146	0.2975
θ	0.0448	0.0244	0.1177	0.1614

Notes: S_L is the standard deviation of the residuals from the ESTAR model for exchange rate i; S_{NL} is the standard deviation of the residuals from the relevant optimised linear VECM; the ratio provides an indication of the improvement in regression fit of explicitly accounting for non-linearity within the real exchange rate with an ESTAR model relative to the basic linear VECM. γ is the estimated smoothness parameter from ESTAR models; θ is the estimated equilibrium correction parameter from the linear VECM. P-values derived from a regression where the optimal ESTAR model has been augmented with the equilibrium correction term from the optimised linear VECM. If the p-value of γ is significant and θ insignificant at standard significance levels, the ESTAR model is said to encompass the linear VECM; and vice versa.

3 Trading the Forward Rate Term Structure

3.1 Introduction

The quality of an exchange rate forecasting model is typically judged by academic researchers on its ability to generate persistently (and significantly) smaller out-of-sample errors than a naive random walk. Using this metric, it seems that little robust progress has been achieved by the academic forecasting community during the two decades that have followed publication of the seminal Meese-Rogoff (1983a, b) papers that found in favour of a random walk over a range of fundamental-based exchange rate models (for a recent survey, see the *Journal of International Economics,* 2003). As we demonstrate in the following chapter, this conclusion seems equally true for forecasting models based upon the emerging microstructural literature as for models based upon more traditional economic fundamentals.

A comparison of out-of-sample forecasting errors derived from theoretical and random walk models is not a particularly useful performance metric in the context of investment portfolio management. Indeed, it represents something of a straw man, a diversion from the principal areas of concern: determining the profitability of investment decisions based upon underlying exchange rate forecasts, and the associated volatility of excess returns. Few studies address these crucial issues in a rigorous manner, while others typically assume either unrealistic-or zero-transaction costs (Rosenberg and Farka, 2001), that investors have perfect foresight (Evans and Lyons, 2002), or that investment portfolios can be turned more frequently than is realistic for most investors other than boutique Hedge Funds or CTAs, given liquidity management issues.

Despite the lack of rigorous academic evidence of an ability to generate exchange rate forecasts that out-perform a naive random walk model, investors have demonstrated a persistent ability to add value to portfolios through currency trading (Baldridge, Meath and Myers, 2000; Hersey and Minnick, 2000). Although these findings appear mutually exclusive, the apparent contradiction is resolved in two ways. First, the quality of academic forecasting models is judged on the size of associated Mean Absolute Forecasting Errors (MAFE) or Root Mean Square Forecasting Errors (RMSFE) relative to a naïve random walk, whereas investors are interested in the profitability of forecasting models irrespective of the size of MAFEs and RMSFEs. Second, academic researchers typically focus upon the accuracy of point exchange rate forecasts, whereas few investors pay these any heed, focusing instead upon the forecast directional path of an exchange rate; persistent forecasting accuracy of this form will achieve investment out-performance relative to an underlying benchmark index as long as the move in the exchange rate is sufficiently large to outweigh associated transaction costs and interest carry.

In this paper, we aim to marry together these two strands of research-academic and investor-using the framework proposed by Clarida and Taylor (1997). Their work-and the subsequent extension by Clarida, Sarno, Taylor and Valente (2003)-represents the first serious contradiction of Meese-Rogoff

(1983a, b) to emerge from academic exchange rate research. It is predicated on the proposition that the forward rate is not an optimal predictor of the future spot exchange rate, but that important information for the future path of the spot rate is nonetheless embedded within the forward rate term structure. Exploiting this information within a linear Vector Equilibrium Correction Mechanism (VECM) estimated by Full Information Maximum Likelihood (FIML), they achieve a statistically significant reduction in forecast errors of the order of 50%-70% relative to a random walk for mark–dollar, yen-dollar, sterling-dollar and French franc-dollar and over forecast horizons that range from 4 to 52 weeks.

Generating forecast errors significantly smaller than a random walk model is certainly an important achievement, but does not guarantee a profitable exchange rate investment strategy. To this end, we replicate the analysis of Clarida and Taylor (1997)-confirming their results in so doing-and then develop a set of trading rules based upon the resulting forecasts that are assessed in terms of their ability to generate returns persistently in excess of a strategic benchmark return. Returns from each of the exchange rate models are examined individually, but also within an equally-weighted portfolio for evidence of diversification benefits that may result from combining the models in this simple manner. We then consider the merits of stop-loss limits that are designed to truncate the extent of negative returns from any trading strategy. We also consider various portfolio construction techniques regularly applied throughout the investment industry to assess the diversification benefits that derive from combining models into portfolios based upon efficient weights that take account of historical return and risk correlations, as well as drawdown parameters that are central to many risk averse investors in the foreign exchange market. We contrast the results of these rules with a naive Forward Rate Bias (FRB) strategy that is widely utilised throughout the foreign exchange investor community.

The remainder of the paper is organised as follows. In the next section we discuss academic evidence on exchange rate predictability, focusing upon the main explanations for the failure of the Efficient Market Hypothesis (EMH) in the foreign exchange market. We then discuss alternative approaches to exploiting evident exchange rate predictability, focusing upon FRB and the framework proposed by Clarida and Taylor (1997). In a subsequent section we present our empirical analysis, and in a final section we draw some conclusions and provide suggestions for future research.

3.2 Literature Review

3.2.1 Failure of the Efficient Market Hypothesis

The market for foreign exchange is the most liquid financial exchange in the world. Daily trading turnover statistics for the foreign exchange market were provided in the previous chapter, and contrast with more modest turnover in bonds and equities. Investors often associate high liquidity with market efficiency. In the context of foreign exchange this appears to be a false association. The EMH is based upon three, related hypotheses: profit maximisation, rational expectations and risk neutrality. Together, these conditions imply that exchange rates should reflect all available information on a continuous basis. Consequently, the expected foreign exchange gain from holding one currency rather than another must just be offset by the opportunity cost of holding funds in this currency rather than the other (Taylor, 1995). The opportunity to make abnormal profits from foreign exchange trading or speculation should be zero.

Beginning with Meese and Rogoff (1983a, b), and until recently, it has been generally accepted among academic researchers that nominal exchange rates are extremely hard to distinguish empirically from random walks (Mussa, 1984). Although a few studies in the intervening period since the publication of Meese-Rogoff papers have demonstrated favourable out-of-sample performance compared with a naïve random walk forecast (e.g. Finn, 1986; MacDonald and Taylor, 1992), these results have generally proven fragile once applied to alternative exchange rates, sample periods or currency arrangements (Sarno and Taylor, 2002).

By contrast, there is considerable evidence amongst practitioners that persistent profit opportunities exist in the foreign exchange market. This apparent conflict between industry and academic evidence seems to reflect two factors. First, a difference in the type of forecasting undertaken by each group. Academic research typically concentrates upon a comparison of the accuracy of out-of-sample point forecasts derived from random walk and fundamental theory-based models of exchange rate behaviour. By contrast, foreign exchange portfolio managers focus upon a relatively easier performance metric: prediction of exchange rate direction, often relative to a 1- or 3-month forward exchange rate.[22] Using this approach, investors have demonstrated an ability to generate persistent excess returns from foreign exchange trading relative to an underlying benchmark index (Baldridge, Meath and Myers, 2000; Hersey and Minnick, 2000). Second, differences in the evaluation method that each group utilises: academic researchers typically judge the quality of competing exchange rate forecasting models on the basis of MAFEs or RMSFEs; investors have little regard for either measure, and instead focus upon the profitability of exchange rate forecasts. Regardless of differences in approach or evaluation metric, the fact that practitioners are able to generate persistent excess returns indicates that the foreign exchange market is inefficient.

Failure of the Efficient Markets Hypothesis in the context of the foreign exchange market can be seen most easily in the stylised fact of FRB (Kritzman, 1993; Clarke and Kritzman, 1996). Practical observation indicates that forward rates typically over-predict the extent of future moves in spot exchange rates, such that the positive differential in the high interest country is only partially eroded by the subsequent exchange rate depreciation. And in many cases forward rates are actually a perverse predictor of future changes in spot exchange rates, such that investors can profit both from a positive interest differential

[22] Foreign exchange portfolio managers typically implement the majority of tactical currency positions in the forward market. This approach requires no up-front cash commitment from clients, thereby allowing funds to be invested in underlying assets to earn a return in addition to the performance of the foreign exchange strategy.

and the subsequent exchange rate move.

To understand the theoretical relationships underlying FRB, we begin with covered interest parity (CIP). This defines the relationship between the current spot and forward exchange rates. The forward premium or discount (that is, the proportionate difference between the levels of spot and forward exchange rates) is the amount that an investor has to pay at time t to hedge exchange rate risk associated with a contract to receive or deliver foreign currency at time t+1. We can write CIP as

$$F_t\left(1+r_t^*\right)=S_t\left(1+r_t\right) \tag{35}$$

or, to a close approximation

$$\left(F_t-S_t\right)/S_t \approx f_t-s_t \approx r_t-r_t^*. \tag{36}$$

A great deal of evidence demonstrates that CIP holds on a continuous basis, meaning that profitable opportunities to arbitrage between current spot and forward rates do not exist in normal trading conditions once realistic transaction costs have been incorporated (Taylor, 1987, 1989).

Assuming CIP holds, uncovered interest parity (UIP) implies that the forward exchange rate is an unbiased predictor of the future spot exchange rate, such that the expected change in the spot rate will be equal to the size of the forward premium plus a rational expectations error term. If we denote the expected value at time t of the spot exchange rate at time $t+1$ as E_tS_{t+1}, we can write the UIP condition as,

$$E_tS_{t+1}\left(1+r_t^*\right)=S_t\left(1+r_t\right) \tag{37}$$

which is approximately equivalent to

$$E_ts_{t+1}-s_t=r_t-r_t^*. \tag{38}$$

Finally, combining CIP and UIP-equations (37) and (39)-gives

$$E_ts_{t+1}=f_t. \tag{39}$$

Assuming that CIP and UIP jointly hold in all periods would allow investors to infer the entire expected future path of the spot exchange rate (Isard, 1995).

Most empirical studies of the validity of UIP estimate some variant of the following equation

$$\Delta_k s_{t+k}=\alpha+\beta\left(f_t^k-s_t\right)+\eta_{t+k} \tag{40}$$

where f_t^k is the logarithm of the forward exchange rate with a maturity of k periods ahead. Assuming risk neutrality and rational expectations, the estimated coefficient α is expected to be insignificantly different from zero, β unity and the rational expectations error, η_{t+k}, to be independently and identically distributed and orthogonal to information at time t.

Table 1 reports the results of standard UIP regressions over the sample period January 1979 to December 2003 using montly data for spot exchange rates and 1-month forward rates. Consistent with stylised fact, our results represent a contradiction of the UIP theory. For all three exchange rates in our study the sign of β is negative, although Wald tests indicate that this coefficient is only significantly different from zero for yen-dollar and sterling-dollar. Studies often assume that the constant term is zero, leading to the interpretation that a negative β means the forward premium is an incorrect directional predictor of the future spot exchange rate direction. In fact our results suggest that this assumption is only valid for mark-dollar, whereas for the other two exchange rates the estimated constant term is significantly different from zero. Nonetheless, our results are generally consistent with the interpretation that the greater the forward premium over some period k, the less the domestic currency is predicted to appreciate (Sarno and Taylor, 2002).

Although UIP is a mainstay of most empirical macroeconometric modelling (Laxton et.al., 1998; Bank of England, 1999), empirical evidence in its support is relatively scare and typically limited to very high (Lyons and Rose, 1995; Chaboud and Wright, 2003) or very low frequency (Flood and Taylor, 1996) data. Lyons and Rose (1995) examine the UIP condition using intraday exchange rate returns and interest differentials for French franc-mark and lira-mark during crisis periods in the Exchange Rate Mechanism (ERM) of the European Monetary System.[23] Lyons and Rose argue that when an exchange rate is subject to the risk of speculative attack, investors will attach to it a significant probability of devaluation for which they require compensation in order to be persuaded to hold assets denominated in this currency. According to UIP, over longer horizons this compensation will be the interest differential between the two countries. During intraday periods, however, there is no interest payment. The only other source of compensation available is variation in the exchange rate itself. Consequently, those exchange rates for which the probability of a crisis is non-negligible but which do not experience devaluation intraday must appreciate over this period to provide appropriate compensation to investors. Lyons and Rose (1995) find evidence consistent with this hypothesis, and thereby conclude in favour of the existence of UIP for intraday exchange rate data.

Chaboud and Wright (2003) adopt a similar approach to Lyons and Rose (1995). In this case, the authors examine intraday exchange rate returns that straddle the close of trading in the New York sector of the foreign exchange market for four exchange rates over a ten-year period from 1988.[24] Overnight open positions accrue interest. If UIP holds this implies that the exchange rate will experience a jump at the session close to reflect this payment. Otherwise an arbitrage opportunity will exist as investors can gain the interest differential

[23] The ERM was the forerunner in Europe of the adoption in January 1999 of irrevocably fixed exchange rates under Economic and Monetary Union (EMU). It was a system of fixed but adjustable exchange rates between member currencies.

[24] Mark-dollar, yen-dollar, sterling-dollar and Swiss franc-dollar. Close of trading is defined as 1700 New York time, adjusted to reflect Daylight Saving where appropriate.

while being exposed to exchange rate risk for an arbitrarily short period of time (Chaboud and Wright, 2003). The authors test this hypothesis using a number of data windows, ranging from a few hours to one day (that is, close to close data). They find evidence in favour of UIP for very short horizons, with the constant and slope terms from equation (41) above insignificantly different from the expected values of zero and unity, respectively. As the data horizon is extended towards twenty four hours, this favourable evidence deteriorates and UIP no longer holds.

That evidence in favour of UIP decays as Chaboud and Wright (2003) extend their investment horizon from intraday to daily is consistent with the majority of empirical research in this area. In addition, the practical value of Chaboud and Wright's finding may be relatively limited as investment banks–who contribute the largest share of transaction volume in the foreign exchange market–provide trading desks with only limited risk budgets to run overnight positions.

In general, for investment horizons beyond intraday regressions based upon equation (41) report estimated slope coefficients significantly lower than one, and often negative (Froot and Thaler, 1990; Engel, 1996). Empirical studies that conclude the estimated slope coefficient is insignificantly different from zero imply that the forward premium is unrelated to the future rate of depreciation (Bilson, 1981).

Flood and Rose (1994) present evidence against UIP on the basis of pooled daily data for a group of floating exchange rates versus the dollar during the 1980s and 1990s. Consistent with other research, they conclude that the relationship between the forward premium and the expected change in the spot rate is negative and significantly different from the unit value hypothesised by UIP. Although their results for fixed exchange rate data from the Europe Monetary System (EMS) are more supportive of UIP–estimated coefficients are at least positive and significantly differently from zero–coefficients nonetheless all lie somewhere below +1. Consequently, even these, more favourable, results are consistent with the presence of inefficiency within the foreign exchange market. Other studies that report similar conclusions to Flood and Rose (1994) include Froot and Thaler (1990) and Fama (1984).

Some authors have argued that the existence of FRB in the 1970s and 1980s was a market anomaly caused by the adoption of floating exchange rates, and the consequent learning period that followed this decision (Baillie and Bollerslev, 2000). In addition, Baillie and Bollerslev argue that the extent of the bias was overstated due to the low power of statistical tests. Accordingly, evidence against UIP and in favour of FRB should have diminished during the 1990s and onwards. In reality, as Meredith and Ma (2002) conclude, there is little evidence–at least for the major exchange rates–to support this conclusion.

More support exists for UIP in the major exchange rates over longer time horizons. Flood and Taylor (1996) examine evidence in favour of UIP using pooled annual data for 21 countries relative to the US over the sample 1973 to 1992 and for a range of investment horizons. Consistent with most other research, on a one year horizon Flood and Taylor conclude that the estimated value of β is significantly lower than +1, although it is positive. But when these data are averaged over five, ten and twenty year periods estimated β's become insignificantly different from the predicted theoretical value of $+1$; as investment horizons are extended UIP appears to be a more relevant explanation of exchange rate determination.

Empirical research has also suggested differences in the extent of FRB between the Developed and Developing Markets. In particular, Bansal and Dahlquist (1999) find evidence in favour of continuous UIP in the Emerging Markets. This is a puzzling result. Emerging Market exchange rates are typically subject to more investor heterogeneity than major markets, particularly in terms of the existence of private information. In association with the relatively low liquidity that also characterises Emerging Markets, this private information leads to periodic jumps in exchange rates, as investors reassess exposure levels to individual currencies in the light of new information or in response to a reassessment of portfolio risk-return preferences in these markets more generally. Consequently, heterogeneity implies that the problem of FRB should be more significant in the Emerging Markets compared with major exchange rates, and not less.

3.2.2 Why Does Forward Rate Bias Exist?

Why this general failure of UIP exists is not well understood, although there are a number of potential explanations. These include the presence of a non-trivial, and time-varying risk premium, expectational errors, rational learning in the context of incomplete information, and so-called peso problems. We discuss each of these in turn. It should be noted that these competing explanations are not mutually exclusive, and that more than one of these factors is likely to be present in market behaviour at any one time (Marston, 1994; Froot and Frankel, 1986).[25] But these factors do embed significant interpretive differences about the behaviour of foreign exchange investors. For instance, explanations of the systematic failure of UIP that rely principally upon the presence of a time-varying risk premium, as well as rational learning and peso problems, maintain the assumption of market rationality. By contrast, explanations based on the presence of systematic expectational errors imply the presence of at least some market irrationality. For comprehensive surveys of these issues, see Lewis (1994), Taylor (1995) and Sarno and Taylor (2002). Empirically, it is difficult to determine the relative importance of these factors within a particular exchange rate (Froot and Frankel, 1986). We also discuss below the possibility, proposed in particular by Baillie and Bollerslev (2000) that the presence of FRB reflects misspecification in the UIP equation due to differences in the statistical properties of exchange rate returns and forward premia.

Risk Premia A number of studies have argued that systematic deviations from UIP can be attributed in large part to the existence of an unobservable time-varying risk premium. This premium represents the market's anticipated

[25] To reflect this observation, Marston (1994) develops joint Wald tests of the competing explanations of risk aversion and expectational errors based upon uncovered interest rate, real interest rate and purchasing power parity conditions.

excess return to holding foreign currency relative to holding domestic currency (Carlson and Osler, 1999). From equation (39) above, this implies that

$$r_t - r_t^* \equiv E_t s_{t+1} - s_t + \rho_t + \varepsilon_t \tag{41}$$

such that the interest differential–or, equivalently, the forward premium–is equal to the expected change in the spot exchange rate, the risk premium, ρ_t, and ε_t, a rational expectations error term, assumed independently and identically distributed with mean zero and variance σ and to be orthogonal to information available at time t.

The most apparent expression of investor risk aversion in the foreign exchange market is the widespread persistence of Home Bias within investment portfolios (Lewis, 1994). Here, investors maintain a higher exposure to domestic assets compared with foreign alternatives than would be justified by asset allocation analysis under the assumptions of perfect substitutability and risk neutrality. This leads on to a core issue in the definition and measurement of risk aversion. Typically, risk-seeking behaviour is associated with the sale of relatively low risk investments, such as high grade government fixed income bonds, in favour of higher risk assets, such as corporate bonds, equities and Emerging market assets. Perhaps more appropriately, risk appetite can also be defined in terms of the introduction and removal of tactical asset allocation decisions, including exchange rate positions, around an underlying strategic benchmark weighting. In this version, rising risk appetite is associated with the introduction of tactical portfolio positions, regardless of asset classes purchased with these positions. It is potentially a crucial differentiation, as it can imply diametrically opposing investment decisions at times of extreme risk appetite. For instance, investor risk aversion increased sharply in the run-up to the 2003 Iraq war (UBS, 2004). International investors began this period underweight the US dollar relative to benchmark indices (CME, 2004) Consistent with the asset-based definition of risk appetite, investors were expected to further reduce holdings of dollars during subsequent months In reality, they purchased dollars, moving back towards benchmark weightings. The implications of this definitional difference of risk aversion would appear to be a interesting area for future research.

The extent of investor risk aversion varies over time, depending upon portfolio positioning, economic and financial market conditions and a range of sociopolitical factors. Indeed, the sign and magnitude of risk premia appear to vary over time, although evidence suggests that they do so in a stationary manner (Barkoulas, Baum and Chakraborty, 2000).

Numerous techniques have been used to derive measurable risk premia. These including Capital Assessment Pricing Models (CAPM), General Equilibrium models (Sarno and Taylor, 2002; Lewis, 1994), and survey data (Froot and Frankel, 1986). Despite the amount of resources expended in this area, the search for stable empirical models of foreign exchange risk premia has not proved fruitful (Taylor, 1995). In particular, empirical risk premia models typically fail to generate sufficient variation in the risk premium – that is, in predicted returns - to explain the high level of actual exchange rate volatility. In addition, although survey data of foreign exchange market activity typically find in favour of the presence of a risk premium in exchange rate returns, the sign of this premium is often inconsistent with the observed directional contradiction of UIP (Froot and Frankel, 1986). Indeed, when the change in the spot exchange rate expected by survey respondents is regressed on interest differentials, the estimated slope coefficient does not differ significantly from one. This indicates that the average survey response provides no support for the hypothesis that rejection of UIP is due to a time-varying risk premium (Isard, 1995).

Practitioners have also attempted to explain systematic deviations from UIP using estimated risk premia. Most efforts have had no demonstrable success, and consequently no ability to add value to investment portfolios in excess of a benchmark index. Of the more successful, the JPMorgan Chase Liquidity and Credit Premia Index (JPMorgan Securities, 1999) has demonstrated a limited ability to generate excess returns relative to an underlying benchmark index in yen-dollar and Swiss franc-dollar by exploiting an estimate of risk appetite based upon a composite index of market-based risk indicators in association with implementation in portfolios of traditional carry trades.[26] These trades have typically focused upon yen-dollar, particularly during the mid-1990s, and Swiss franc-dollar. They appear to be more profitable when risk appetite is relatively high. Consequently, a combination of positive carry (that is, low Japanese or Swiss interest rates relative, say, to US rates) and high investor risk appetite implies that these trades can profitably be implemented in investment portfolios. Alternatively, when the LCPI indicates that investors are risk-averse yen-dollar and Swiss franc-dollar carry trades should be unwound and foreign exchange positions returned to the neutral, or underlying strategic, benchmark. But profitable trades associated with the prediction of the LCPI have been achieved over a limited sample period, and the quality of information appears to have deteriorated in recent years suggesting excessive fragility. Furthermore, this tool has achieved little success with other currency pairs (JPMorgan Fleming Asset Management, 2002). Emphasising this fragility, this index has recently been altered to incorporate a measure of option volatility, although it is not clear that this change has led to an improvement in performance.

Forecasting Errors The second possible explanation for the presence of systematic deviations from UIP reflects the presence of persistent forecasting errors. These errors imply a failure of the Rational Expectations Hypothesis (REH). Froot and Frankel (1986), on the basis of survey evidence, conclude that the presence of persistent forecasting errors is the principal source of market inefficiency.

Hence,

[26] A carry trade involves borrowing funds in low interest rate currencies and investing them in higher rate currencies. Investors expect to be paid both the positive interest differential and by an appreciation of the exchange rate.

$$r_t - r_t^* = E_t s_{t+1} - s_t + \varepsilon_t \tag{42}$$

where $E(\varepsilon_t)$ no longer equals zero, but instead includes a systematic, predictable component over time. Following Evans and Lewis (1990), we note that this appearance of irrationality in forecasts may be visible only on an ex post basis, and that investors may actually be using information in an ex ante rational manner. For this reason, Evans and Lewis (1990) term these forecast errors biased, and not irrational.

Systematic forecasting errors may reflect the presence in the foreign exchange market of three factors. First, investors may place excessive weight upon new information relative to the current spot rate when forming forecasts of the future exchange rate (Froot and Frankel, 1986). Second, it is estimated that only a small fraction of average daily foreign exchange flows are initiated by profit-maximising investors, typically in currency overlay firms, Hedge Funds, CTAs and on proprietary trading desks of investment banks. Most flows are not inspired by a profit motive, but instead reflect non-level dependent hedging activity by corporate treasurers and passive investors that will tend to chase the spot exchange rate, or foreign exchange intervention by central banks designed to ensure orderly market conditions or achievement of longer-term monetary policy goals such as inflation control or growth maximisation. And third, the presence of technical, or feedback, investors that introduces trends into exchange rates that are not justified by fundamentals, but may nonetheless generate returns to this activity in excess of an underlying benchmark return. Indeed, despite this traditional classification, the extent to which technical investors have demonstrated the ability to generate persistent excess returns suggests that they are in fact acting rationally and exploiting the proven unit root properties of spot exchange rates in the vicinity of equilibrium (Taylor, Peel and Sarno, 2001).

Rational Learning and Peso Problems Systematic departures from UIP may also reflect the existence within the foreign exchange market of either rational learning or so-called peso problems. The two are closely related. Rational learning appears to be consistent with the marked information asymmetries that exist in the foreign exchange market, for instance relating to the wide variety of market participants, as well as the incidence of central bank intervention and the timing and magnitude of large investor asset allocation shifts between domestic and international assets and between equities and bonds.[27]

Rational learning occurs following a shock, typically characterised as a fundamental shift in policy stance. In its aftermath, investors take time to learn about the new policy environment and, consequently, attach a significant probability of another policy shock occurring in the future. During the period that investors learn that the new policy environment is in fact permanent forecast errors will exhibit significant serial correlation, implying that expected changes

[27] As the hedge ratio on international equities in investment portfolios is typically lower than for foreign bonds, allocational shifts between these asset classes will often have implications for exchange rates.

in exchange rates will not equal interest differentials and therefore that UIP will not hold. However, it is clear that this explanation, as well as the peso problem discussed below, can only explain periodic failure of UIP; agents cannot forever learn about a once-and-for-all regime shift (Sarno and Taylor, 2002) just as once-and-for-all regime shifts, by definition, do not happen all the time.

Alternatively, we may assume that investors learn instantaneously about the price implications of a fundamental policy shift, but remain convinced that another jump shift in policy will occur in the future. This issue is termed a peso problem, following the behaviour of the Mexican peso in the 1970s, whereby it sold at a forward discount for an extended period ahead of its eventual devaluation in 1976 (Rogoff, 1979). On a more general level, the term "peso problem" has been used to describe a non-negligible probability of a shift between appreciating and depreciating periods (Engel and Hamilton, 1990), or regimes (Evans and Lewis, 1992) over some fixed investment horizon.

Following Evans and Lewis (1992) we can define the expected exchange rate at time t+1 in the presence of a peso problem as

$$E_t s_{t+1} = (1-\beta) E_t (s_{t+1} \mid C) + \beta E_t (s_{t+1} \mid A) \tag{43}$$

where $(s_{t+1} \mid C)$ is the expected exchange rate based upon the current generating process generating the exchange rate, $(s_{t+1} \mid A)$ is the expected exchange rate based upon some alternative generating process and β is the probability of a switch from the current to the alternative process.

In the presence of a peso problem expected exchange rate changes will diverge from interest differentials as a result of serially correlated, but entirely rational, forecast errors. This happens as investors price in a significant probability of a future exchange rate devaluation over some finite horizon, and thereby demand higher interest payments on the depreciating currency in compensation for holding assets denominated in it. In reality, the devaluation does not occur during this period, but does take place at some later time, thereby vindicating market expectations. In retrospect, though, the interest differential was excessive given the change in the exchange rate that actually occurred during that period. Examples of peso problems have been cited by Evans and Lewis (1991), Lewis (1991), Clarida and Taylor (1993), and Kaminsky (1993).

The discussion of the failure of UIP has concentrated so far upon factors that drive a systematic wedge between expected exchange rate returns and forward premia. Accounting for these factors in a robust manner could in principle restore the theoretical UIP relationship. The final explanation for the failure of this relationship, due primarily to Baillie and Bollerslev (2000), emphasises the lack of any such theoretical underpinning. This follows from the statistical properties of the two components of UIP. Baillie and Bollerslev find that exchange rate returns are distributed as a stationary series – that is, $I(0)$ - with a constant mean and variance. By contrast, they conclude that the forward premium is a fractionally integrated, or long memory series, that is $I(d)$, where $0 \prec d \prec 1$. Maynard and Phillips (2001) reach similar conclusions using both semiparametric and parametric estimation methods, and Crowder (1994) concludes on the basis of Augmented Dickey Fuller tests for the mark and Canadian dollar expressed in terms of the US dollar using monthly data over the sample period January 1974 to December 1991 that the forward premium is actually a unit root series, or $I(1)$.[28] All of these results imply that shocks to forward premia are much more persistent than equivalent shocks to exchange rate returns, and that these two variables cannot exhibit a long-term cointegrating relationship. Consequently, UIP regressions are spurious and any inference drawn on the basis of this relationship is meaningless. Our results, reported below, contradict those of Baillie and Bollerslev (2000). Instead we find clear evidence that forward premia are stationary for three exchange rates–mark-dollar, yen-dollar and sterling-dollar–and four different premia incorporating 1-, 3-, 6- and 12-month forward rates. As our data spans twenty four years the inference we draw should be more robust than other studies for which data samples are rather shorter. Confirmation for our findings comes from Clarida and Taylor (1997), using essentially the same data set, and Barkoulas, Baum and Chakraborty (2000), who use a data panel for a similar range of exchange rates and forward premia and over a similar time span. In addition, Granger (1999) suggests that the results of Baillie and Bollerslev (2000) are themselves spurious, and caused by the presence of regime shifts or structural breaks in the forward premia, both of which may reflect the presence of other factors discussed above.

3.2.3 Can Investors Exploit the Failure of UIP?

Although an unequivocal understanding of the reasons why UIP fails on a systematic basis has alluded researchers and practitioners alike, it seems clear that market inefficiency is present within the foreign exchange market. The next issue to investigate is whether this inefficiency is actually sufficiently large and persistent over time for investors to exploit it in order to generate returns within investment portfolios in excess of underlying benchmark indices, including transaction costs of implementing the associated trading strategy. Consideration must also be given to the strategic impact of trading activity upon market prices, which in turn has implications for the magnitude of feasible position sizes The results of this analysis will help determine the applicability of the infamous Meese-Rogoff (1983a,b) results to a practical portfolio investment context.

There are potentially a number of ways to exploit this failure of UIP. We concentrate upon two in this paper. First, a relatively naïve investment strategy based directly upon FRB trades. Second, development of a disciplined set of trading rules based upon information extracted from the forward exchange rate term structure. This information is extracted using the approach of Clarida and Taylor (1997). The accuracy of point forecasts derived from the Clarida-Taylor framework will be compared with a naïve random walk strategy using MAFEs and RMSFEs, as well as the Diebold Mariano (DM; Diebold and Mariano, 1995) test statistic for equality of forecast accuracy. The DM test statistic can accom-

[28]This result seems a little implausible, given the arbitrage opportunities that it implies. Crowder (1994) also reports ADF tests on sterling-dollar forward premia data over the sample period; these tests reject the null of a unit root.

modate a wide variety of features relating to forecast errors, including non-zero means, non-normality and contemporaneous correlation between rival prediction methods applied to the same time series. As such, the DM test statistic is a particularly flexible and appealing metric by which to assess forecast accuracy across various prediction methodologies. The DM test statistic is defined as:

$$DM = \frac{\bar{d}}{\sqrt{\frac{2\pi \widehat{f}(0)}{N}}} \tag{44}$$

where $\bar{d}$ is an average over N forecast periods of a general loss function d_x such as the difference in squared forecast errors

$$\left(d_k = \left[\zeta_k^{(1)}\right]^2 - \left[\zeta_k^{(2)}\right]^2\right) \tag{45}$$

or in absolute errors

$$\left(d_t = \mid \zeta_k^{(1)} \mid - \mid \zeta_k^{(2)} \mid\right), \tag{46}$$

that is $\bar{d} = \frac{1}{N}\sum_{k=1}^{N} d_k$ for $N = 1,..,42$ in our recursive application; $\widehat{f}\,(0)$ is a consistent estimate of the spectral density of the loss differential function at frequency zero, which we estimated using the method of Newey and West (1987). Under certain regularity conditions the DM test statistic will be distributed as standard normal under the null hypothesis of equal forecast accuracy (Diebold and Mariano, 1995).[29]

In the case of a random walk prediction, the optimal one-period ahead exchange rate forecast is simply today's value with perhaps also a constant drift component; this represents the traditional academic metric against which the performance of theory-based models is assessed. In addition, a set of trading strategies will be constructed to test the profitability of both approaches, incorporating realistic transaction costs, feasible trade sizes and appropriate portfolio restrictions. Profitability is calculated, initially, in terms of the Information Ratio of a strategy, which is defined as a measure of excess return per unit of risk taken, where excess return is defined as the difference between the currency portfolio's return and the return to a benchmark index,

$$IR = \frac{E(R_{s)} - R_{rf}}{\sigma_s} \tag{47}$$

where $E(R_{s)}$ is the expected return to the trading strategy-proxied by historical return $E(R_{s)}$-R_{rf} is the return to the risk-free interest rate, or equivalently the strategic portfolio benchmark-and σ_s is the standard deviation of the active returns. $E(R_{s)} - R_{rf}$ is often termed the excess return accruing to an active

[29]The distribution of the DM test statistic is unclear in small samples. Consequently, marginal significance levels reported below should be interpreted with caution.

trading strategy. In addition, we will also examine the profile of cumulative returns to trading strategies, with particular focus upon the length and depth of capital drawdown periods. Taken together, these metrics will allow us to consider the consistency of trading strategies with the performance objectives and time horizons of a typical foreign exchange investor, for instance a corporate pension fund or a sovereign debt manager.

Overall, therefore, our analysis will provide a rigorous examination of the ability of foreign exchange trading strategies based upon fundamental theoretical approaches to generate persistent returns in excess of a benchmark return, and in line with wider portfolio performance objectives than extant work in this area, for instance see Bilson (1981).

Forward Rate Bias[30] FRB is closely related to the carry trade notion discussed above. Although FRB strategies can involve any exchange rates, or combinations of rates, they typically involve constructing a basket of currencies against the dollar (Rosenberg and Farka, 2001; Strange, 1998). Rosenberg and Farka (2001) use daily data from January 1986 to December 2001 to construct an equally weighted basket of ten currencies against the dollar.[31] Long positions are established in those currencies where the 3-month euro deposit interest rate exceeds the equivalent US rate, that is in those currencies that trade at a forward discount to the dollar. Similarly, short positions are implemented for those currencies trading at a forward premium to the dollar. Positions are opened and closed at the beginning and end of each month, and the underlying strategic benchmark return is assumed to be US cash. Alternatively, one can calculate a net interest rate based on the basket and then compare this with the US dollar interest rate (Strange, 1998). For a net positive premium in favour of the basket, this strategy would short the dollar; conversely, for a net interest rate discount on the basket, the strategy would recommend long dollar positions relative to the benchmark position. As before, these positions are typically reset every period, for instance monthly. Although slightly different in design, both strategies deliver a binary outcome – portfolios are either long or short the dollar in every period, with no neutral active currency allocation.

FRB strategies can also incorporate no-arbitrage thresholds, consistent with the non-linear Threshold AutoRegressive (TAR) models of Tong (1990). In this case, trades are not introduced into portfolios for small non-zero interest differentials, as the benefit of arbitrage from these positions is likely to be outweighed by the cost of implementation. Clearly, following on from the discussion of the previous section, the presence of a non-zero risk premium can imply the existence of interest rate thresholds that prohibit arbitrage of small differentials. As the size of this premium can differ between domestic investors in different countries, these thresholds are likely to be asymmetric for any given exchange rate.

[30] This strategy is also often termed Forward Discount Bias. The two names are synonymous.

[31] The euro, yen, Swiss franc, Canadian dollar, Australian dollar, New Zealand dollar, Danish krone, Norwegian krone and Swedish krona.

Although simple to calculate and implement, performance metrics for FRB trading strategies are typically unimpressive once due allowance has been made for realistic transaction costs and feasible position sizes. Without transaction costs, and relative to US cash, Rosenberg and Farka (2001) report annual returns in excess of US cash to their FRB strategy of 5.64%, with an Information Ratio (IR) of 0.73. This ratio is high compared with the long-term investment performance reported for other asset classes, including equities (Lyons, 2001). But it is important to incorporate realistic transaction costs within any trading strategy. On a notional US$100 million position reset at the beginning of each month, and assuming normal market trading conditions, realistic round trip transaction costs range from 5 basis points for euro-dollar and yen-dollar, to 25bps for Australian dollar-dollar and 100bps for New Zealand dollar-dollar.[32] Applying these costs to Rosenberg and Farka (2001), the IR of this strategy falls to 0.30, with annual average excess returns of 2.78%. By implication, this strategy will introduce almost 10% risk into investment portfolios, which is likely to far exceed tolerable limits for most foreign exchange investors. For instance, a typical corporate currency overlay client will accept an annual average risk budget for this strategy in the region of 2.5% (JPMorgan Fleming Asset Management, 2003a). Alternatively, therefore, imposing this tracking error limit the annual excess returns to this FRB strategy fall to around 0.75%. Given typical management fees (JPMorgan Fleming Asset Management, 2003a), this is a much less compelling result in favour of naïve FRB trading strategies than initial results implied. In addition, the performance of these rules appears relatively lacklustre compared with more sophisticated trading strategies that rely on a combination of fundamental economic, market microstructure, technical and qualitative inputs, and that may also incorporate relatively rigorous econometric techniques. For instance, Baldridge, Meath and Myers (2000) report that the universe of currency overlay managers generated an average IR of 0.55 over the period 1989-2000, using various strategies along these lines; performance results reported by Hersey and Minnick (2000) provide a similar conclusion.

Periods of under-performance associated with FRB strategies are also typically persistent and deep. This is a particularly important consideration for risk averse investors with a particular sensitivity to capital losses, for instance central bank foreign exchange reserve managers. Using the FRB strategy of Rosenberg and Farka (2001), the longest underperformance period persists for more than four years, during which time the maximum peak-to-trough capital drawdown exceeds 18%.[33] Most investment plan sponsors operate with an horizon of about 2 years; some impose explicit stop-loss or Value-at-Risk constraints upon portfolio managers, typically well short of 5% of the capital value of the

[32] On a notional $100mn transaction, we assume the following transaction costs: 5 basis points (bps) for the euro, yen, sterling and Danish krone; 10bps for the Canadian dollar and Swiss franc; 15bps for the Norwegian krone and Swedish krona; 25bps for the Australian dollar; and 100bps for the New Zealand dollar.

[33] Longest Underperformance is the longest period of time it takes cumulative returns from a trading strategy to return to its previous local peak after a drawdown period. Maximum peak-to-trough drawdown is the largest capital loss suffered by a trading strategy from the local high point of cumulative returns to the local low.

total investment portfolio. Consequently, it appears clear that FRB strategies, although useful for crystallising the concept of foreign exchange market inefficiency, are inappropriate as direct trading strategies.

Modelling the Forward Rate Term Structure The performance and use of FRB trading strategies is well documented in the literature. An alternative, and more sophisticated, approach to exploiting foreign exchange market inefficiency is presented by Clarida and Taylor (1997) and developed further by Clarida, Sarno, Taylor, and Valente (2003). Although systematic failure of UIP indicates that the forward rate at any maturity is a biased predictor of the future spot exchange rate, Clarida and Taylor develop an empirical framework that is able to accommodate-indeed, is agnostic about the reasons for-rejection of the pure efficiency hypothesis while still allowing forward premia to contain important information about future spot exchange rate changes (Sarno and Taylor, 2002).

The Clarida-Taylor framework determines and exploits the presence of $n-1$ long-term cointegrating equilibria between spot exchange rates for the mark, yen, sterling and French franc, expressed in terms of the dollar, and associated forward exchange rates at 4-, 13-, 26- and 52-week maturities. The existence of these $n-1$ cointegrating equlibria for each exchange rate implies that these series demonstrate long-term co-movement, and that a predictable dependent relationship exists between them. Consistent with the Granger Representation Theorem (Granger, 1983), this predictable relationship can be exploited within a system of linear VECMs, with forward premia $(f_t^k - s_t)$ acting as the equilibrium correction terms; these terms ensure that the relationship between spot exchange rates and each of the forward rates returns to equilibrium in the wake of an unanticipated shock.

To derive the Clarida-Taylor framework, and using the results of Beveridge and Nelson (1981), we can decompose the spot exchange rate into two components,

$$s_t = m_t + q_t \tag{48}$$

where m_t is a random walk with drift component and q_t is a zero mean stationary component. If we define departures from the EMH as

$$\gamma_t \equiv f_t^k - E(s_{t+k} \mid \Omega_t) \tag{49}$$

combining equations (11) and (12) gives

$$f_t^k = \gamma_t + k\theta + E(q_{t+k} \mid \Omega_t) + m_t \tag{50}$$

where θ is the drift of the random walk component, m_t. Subtracting equation (11) from (12) gives

$$f_t^k - s_t = \gamma_t + k\theta + E(q_{t+k} - q_t \mid \Omega_t). \tag{51}$$

By definition, θ is a constant value and $E(q_{t+k} - q_t \mid \Omega_t)$ is a stationary, or $I(0)$, series. Accordingly, $f_t^k - s_t$ will also be stationary as long as γ_t, deviations from the EMH, is $I(0)$. If this assumption holds, there exist $n-1$ cointegrating equilibria between f_t^k and s_t; these equilibria represent the core of the Clarida-Taylor approach.

Following Clarida and Taylor (1997), and given the existence of these cointegrating equilibria, we can write the linear VECM between spot and forward exchange rates as

$$\Delta y_t = \eta + \Sigma_{i=1}^{j} \Gamma \Delta y_t + \Pi y_{t-1} + \varepsilon_t \tag{52}$$

where η is a deterministic coefficient, $y_t = \left(s_t, f_t^4, f_t^{13}, f_t^{26}, f_t^{52}\right)$ is the system's $j+1=5$ by 1 vector of variables for a particular exchange rate. Πy_{t-1} is a vector of levels equilibrium correction terms that incorporates the four forward premia, $\left(f_t^4 - s_t, f_t^{13} - s_t, f_t^{26} - s_t, f_t^{52} - s_t\right)$. The speed of reversion towards equilibrium in the wake of an unanticipated shock is therefore determined by the coefficient vector Π. Finally, ε_t is an error term assumed independently and identically distributed with mean zero and variance σ. Clarida and Taylor estimate the VECM system for each exchange rate by FIML over the period January 1979 to December 1995, using weekly data. VECM systems are optimised using a traditional general-to-specific procedure, with the least significant coefficient removed from the system at every iteration until all coefficients are significant at a 5% level. Clarida and Taylor then generate a series of out-of-sample forecasts over the period January 1996 - December 1998 based upon the optimised version of this system for each exchange rate, with forecast horizons ranging from 1 to 52 weeks. Consequently, these forecasts only use information publicly available at the time that they are prepared; they contain no perfect foresight. Clarida and Taylor assess the performance of these VECM Term Structure exchange rate forecasts against a naïve random walk forecast, where the n-period ahead optimal forecast is simply today's spot rate, on the basis of a comparison of MAFE and RMSFE statistics, as well as the DM test of forecast accuracy.

Adopting this VECM approach, Clarida and Taylor achieve a significant improvement in forecast accuracy compared with either a naive random walk at almost forecast horizons. This improvement indicates that the entire term structure of forward premia out to one year contains statistically significant information about the future path of the spot exchange rate that is not embedded in the lagged change of the spot rate itself (Clarida and Taylor, 1997). The general exception is at a one-week horizon, where the Term Structure models achieve similar results to a naïve random walk. This is unsurprising, as most evidence suggests that for high frequencies nominal exchange rates exhibit a unit root, and are therefore distributed as a random walk (Mussa, 1984). For longer forecast horizons the reduction in MAFEs and RMSFEs relative to the naïve random walk is stark. For mark-dollar, the improvement ranges from 14% at 10-weeks, to 33% at a 26-week horizon and almost 50% at a 52-week horizon. For sterling-dollar, the improvement in RMSFEs reaches 90% at 52 weeks. This is an impressive achievement indeed. By way of confirmation, DM test statistics indicate that this improvement in forecast accuracy in favour of the Clarida-Taylor framework is significant at the 1% level.

Clarida, Sarno, Taylor and Valente (2003) further develop this VECM framework to exploit nonlinearity inherent in the relationship between spot and forward rates. They do this by developing a three regime Markov-Switching VECM. This augmentation leads to a further improvement in out-of-sample forecast accuracy. In this case, the average gain relative to the linear VECMS, across each of the exchange rates and performance criteria discussed above, is 1%-10% at 4-weeks and 10%-30% at 52 weeks. DM test statistics again indicate that these improvements are statistically significant.

The predictive accuracy of the Clarida-Taylor linear VECM system, and the Clarida-Sarno-Taylor-Valente non-linear augmentation, is outstanding. Although other researchers during the past twenty years have achieved some out-of-sample forecasting success relative to a naïve random walk model using fundamental based models (e.g. Finn, 1986; MacDonald and Taylor, 1992), these results have typically proved to be fragile, and specific either to time period or exchange rate arrangement, or both. By contrast, the Clarida-Taylor results arguably represent the first robust contradiction of the Meese-Rogoff (1983a, b) stylised fact that fundamental-based out-of-sample point exchange rate forecasts are persistently inferior to a naïve random walk.

The implications of this result for the active management of foreign exchange exposure in investment portfolios are not immediately clear. Achieving a significant reduction in RMSFEs compared with a naïve random walk has no direct mapping to the performance of an investment portfolio that incorporates substantial exposure to international assets and exchange rates. Consequently, we apply the Clarida-Taylor methodology in an investment context and assess whether this approach could be used to generate persistent excess returns relative to an underlying benchmark index return by active management of foreign exchange portfolio exposure.[34] In order to make this assessment we apply a set of trading rules to out-of-sample forecasts for each of our three exchange rates over the period January 1999 to December 2003, using a range of investment portfolio criteria to determine the profitability of these trading rules, as well as the extent to which the volatility and drawdown characteristics of returns are consistent with the risk-return objectives of a typical risk-averse international investor.

The most common investment performance criterion is the Information Ratio (IR), as defined above. IR calculations assume that investors are indifferent to the drawdown characteristics of returns. This is rarely the case, and many clients introduce explicit calendar or rolling year stop-loss limits into portfolios that require currency managers to remove active hedges if these limit levels

[34] Although the non-linear augmentation to the original Clarida-Taylor linear VECM system generates a statistically significant improvement in predictive accuracy of the future path of spot exchange rates relative to the linear alternative it is computationally difficult to apply in a practical investment portfolio context. Consequently, we content ourselves with an assessment of the original linear VECM framework.

are breached. Accordingly, we also report a series of performance measures that incorporate aspects of downside performance. First, Maximum Drawdown measures the largest peak-to-trough decline in cumulative returns during the simulation period. Related, we report sterling ratio data that compare the level of annualised returns to the maximum drawdown, thereby providing an indication of the expected payoff to the simulated strategy, as well as the 95% confidence interval of annualised returns that provides an indication of the inherent skill of the trading strategy. Second, Longest Underperformance indicates the number of weeks it takes cumulative returns to recover the level of the previous local peak in the aftermath of a capital drawdown. Third, sortino ratios relate annualised returns to semi-standard deviation, or downside risk. This measure calculates the level of risk introduced by a trading strategy into portfolios using only the volatility of negative returns. This focus exclusively upon undesirable returns appears to be an improvement upon IR. However, its use is much less common than IR in investment management, perhaps reflecting the fact that the statistical properties of semi-standard deviation are more ambiguous than standard deviation. In addition, as return asymmetries may not be stable over time, using realised downside risk may not be a satisfactory proxy for future downside risk. Finally, to the extent that returns are symmetric around zero the semi-standard deviation will simply be proportional to standard deviation and its calculation will add no insight (Grinold and Kahn, 1999). Nonetheless, in many cases returns to active currency hedges are not symmetric around zero, and so-interpreted with caution-the sortino ratio can provide additional insight into the performance of simulated trading strategies, and the Clarida-Taylor framework.

3.3 Empirical Results

3.3.1 Data

Our empirical analysis uses weekly data for mark-dollar, yen-dollar and sterling-dollar, all expressed as the domestic price of foreign currency, and associated forward rates out to one year over the period January 1979 to December 2003. Data until December 1998 is used for in-sample optimisation of linear VECMs, with data from January 1999 onwards retained for out-of-sample forecasting.[35] Our data set therefore includes 1304 observations in total, with 261 of these observations out-of-sample. Data are taken from the Bank for International Settlements (BIS) and DataStream. Forward premia are defined to be consistent with equation (37) above.

3.3.2 Clarida-Taylor Term Structure Models

Our first step in this section is to verify the results of Clarida and Taylor (1997). We then develop a set of profitable trading strategies based upon these results,

[35] Although included in the Clarida-Taylor analysis, we do not analyse the French franc-dollar exchange rate in this study, given its replacement in 1999 by the euro-dollar rate.

thereby marrying together academic and investor strands of exchange rate forecasting.

If information useful to the prediction of spot exchange rates does exist within the forward rate term structure then each of the forward rates used in the Clarida–Taylor framework-4, 13, 26 and 52 weeks-should exhibit similar statistical properties to the associated spot exchange rate, and a unique cointegrating relationship will exist between spot and each forward rate. Within the VECM system, this implies that there are $n-1$, or 4 cointegrating vectors for each exchange rate system.

We test for the presence of a unit root in spot and forward rates using the Augmented Dickey Fuller (ADF) test. For a discussion and definition of the ADF test, see the previous chapter. Lag length for ADF tests is chosen to minimise the Akaike Information Criterion (AIC) in order to ensure against the presence of neglected serial correlation that may otherwise bias the results of these tests. Table 2 reports the results of ADF tests. These indicate that each of the spot exchange and forward rates under examination contains a single unit root. This is consistent with the findings of Clarida and Taylor (1997), as well as Barkoulas, Baum and Chakraborty (2000) based upon a panel data study of six major exchange rates.

Evidence of a single unit root for each series in our data set represents a necessary but not sufficient condition for the existence of a set of long-term equilibria between each of the three exchange rates and associated forward rates. A crucial foundation of the Clarida-Taylor approach is that these equilibria exist. To test the validity of this assumption we use the Johansen Maximum Likelihood cointegration test. The results are presented in Table 3. Consistent with a priori expectations and the findings of Clarida-Taylor, Johansen tests for yen-dollar and sterling-dollar indicate the presence of four cointegrating equilibria between spot and forward rates; forward premia for these exchange rates are stationary series. This makes intuitive sense, and implies that premia will tend to revert back to an equilibrium value–which may be zero-in the wake of an unanticipated shock.

By contrast, the results for mark-dollar appear a little unintuitive. Johansen cointegration tests indicate the presence of only three cointegrating vectors between spot and forward rates.[36] It is unclear why this result should occur, and given pervasive evidence in favour of CIP beyond an initial transaction cost threshold (for surveys, see Taylor, 1987, 1989), it is difficult to perceive of a situation in which a persistent and growing differential could emerge between mark-dollar and forward rates of any maturity; this exchange rate (now euro-dollar) accounts for 37% of total daily trading volumes in the foreign exchange market (BIS, 2004), suggesting that profitable trading opportunities caused by a divergence in spot and forward rates will be arbitraged by investors very quickly.

As there appears to be no valid economic explanation for this result, it may instead reflect the low power of the Johansen test in the context of a relatively

[36] This result contrasts with the findings of Clarida and Taylor. The reason for this difference is not known.

short data span such as the one employed within our study. This power problem reflects the long-term nature of the hypothesis embedded in the test: with a short data span, regardless of the frequency of observations, it can be difficult to distinguish a mean-reverting series from a random walk, and therefore to reject the null hypothesis of no cointegration even there does in reality exist some linear combination variables that forms a cointegrating vector.[37] Consequently, we proceed on the basis that there exist four cointegrating vectors for mark-dollar, as well as yen-dollar and sterling-dollar.

The results of overidentifying restriction tests on the β' matrix of cointegrating coefficients generally support this conclusion. These restrictions are imposed to uniquely identify the cointegrating vectors and take the form,

$$\beta' y_t = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_t \\ f_t^4 \\ f_t^{13} \\ f_t^{26} \\ f_t^{52} \end{bmatrix} \tag{53}$$

If the relationship between the various spot exchange rates and each forward rate was exactly in line with theoretical priors at all times, these coefficient restrictions would be accepted by the data; the estimated coefficients of a forward premium at any maturity k would be equal to $[1, -1]$ on a continuous basis. As the results in Table 4 indicate, this is not the case for any of the three exchange rates in this study and the restrictions are rejected, at a 5% significance level. To determine the magnitude of the departures from a strict spot-forward relationship, we re-run these tests, but now impose the following exactly identifying restrictions,

$$\beta' y_t = \begin{bmatrix} -1 & \phi_4 & 0 & 0 & 0 \\ -1 & 0 & \phi_{13} & 0 & 0 \\ -1 & 0 & 0 & \phi_{26} & 0 \\ -1 & 0 & 0 & 0 & \phi_{52} \end{bmatrix} \begin{bmatrix} s_t \\ f_t^4 \\ f_t^{13} \\ f_t^{26} \\ f_t^{52} \end{bmatrix} \tag{54}$$

The results of these tests are reported in Table 5 and indicate that departures from the overidentifying restrictions are relatively small: estimated coefficients for yen-dollar and sterling-dollar are very close to theoretical values; departures for mark-dollar are a little larger, but still relatively small and may perhaps be explained by tiny data imperfections (Clarida, Sarno, Taylor and Valente, 2003). Overall, therefore, given the body of theoretical and practical evidence in favour of the existence of four cointegrating vectors between spot and forward exchange rates for each of the three exchange rates in our study-as well as the general thrust of our results-we feel it reasonable to conclude in favour of the Clarida-Taylor test results.

[37] Other explanations may include the presence of heteroscedasticity in the cointegrating relationships.

Consistent with the Granger Representation Theorem (Granger, 1983), the existence of a set of four cointegrating equilibria between mark-dollar, yen-dollar and sterling-dollar and associated forward rates implies the presence also of a dynamic characterisation of the short-term relationship between these variables. Both aspects of this relationship can be captured efficiently within a VECM. In the Clarida-Taylor framework, it is expected that significant interdependencies exist between each equation within individual VECM systems–within each system, there are five equations, one with the spot exchange rate as the dependent variable and one for each of the four forward rates. The best way to exploit these interdependencies is to estimate the VECM by FIML. This procedure estimates the likelihood function under the assumption that the contemporaneous errors within the VECM have a joint normal distribution. Provided that the likelihood function is correctly specified FIML is fully efficient. We select the optimal lag length for each VECM system using a Wald Likelihood Ratio test; this test assesses the joint significance of all i-th lagged endogenous variables in the VECM system, and has a chi-square distribution with k^2 degrees of freedom under the exclusion null. We then undertake a traditional general-to-specific procedure to sequentially eliminate insignificant coefficients within the system until we achieve a set of parsimonious VECMs within which all remaining coefficients-including equilibrium correction terms-are significant at a 5% level. These parsimonious VECMs are presented in Table 6.

The optimised VECM systems are then used to generate a series of n-step ahead dynamic forecasts, for each of the five dependent variables in our three systems.[38] As we are only interested in the forecasts of the spot exchange rates, we report only the associated performance statistics-the ratio of MAFEs and RMSFEs for the VECM out-of-sample forecasts relative to a naïve random walk forecast-for these series, in Tables 7, 9 and 11. Our results are consistent with the findings of Clarida and Taylor (1997). Across all forecast horizons, except one week for sterling-dollar, Term Structure models achieve much greater forecasting accuracy than a naive random walk. At a 4-week horizon, this improvement is equivalent to a 30%-50% decline in forecast errors, with this improvement increasing to 60%-80% at a 52-week horizon. These improvements are statistically significant, as indicated by DM test statistics reported in Tables 8, 10 and 12.

The results of our analysis so far have supported the theoretical foundations on which the Clarida-Taylor approach is predicated, and also the marked improvement that it achieves relative to the standard academic metric of a naive random walk. This in itself is a major achievement compared with the existing literature on exchange rate forecasting (*Journal of International Economics*, 2003). But these results indicate little about the ability of this approach to generate persistent profits in an investment portfolio context. Consequently,

[38] I thank Giorigio Valente for the use of his Ox programme in generating these forecasts. Consistent with Clarida and Taylor (1997), VECM systems used to generate MAFEs and RMSFEs incorporate only one lag for each variable; VECMs employed to generate out-of-sample forecasts used in portfolio simulations below adopt a richer lag structure based upon the selection criteria detailed here.

the next step in our assessment of the Clarida-Taylor approach is to determine whether an improvement in forecast accuracy as measured by MAFEs and RMSFEs can translate into persistent value-added in excess of a strategic benchmark return. To make this assessment, we run a series of simulations, introducing active currency hedges into a representative portfolio around a benchmark currency exposure. The sign of these active hedges will be determined by weekly directional forecasts generated by the Clarida-Taylor approach; by contrast, the MAFE and RMSFE data reported above were derived from point exchange rate forecasts. Assumed transaction costs-at 5 basis points round trip-are consistent with the notional 5% position size assumed throughout this section.[39] To ensure that simulation results do not incorporate perfect foresight we include a one day lag between the production of weekly forecasts and implementation of associated active currency positions into portfolios: forecasts are generated using WM/Reuters (WMR) closing exchange rates on day t, and implemented into portfolios at WMR closing rates on day t+1.[40] All performance data are simulated over the sample January 8, 1999 to December, 26 2003.

Our initial simulation is based upon a simple buy-sell trading rule. We implement a long 5% position in currency i into portfolios whenever the 4-week forecast level of spot exceeds the level of the 4-week forward rate; similarly, a short position is implemented whenever the directional forecast of spot is below the forward rate; there is no neutral position.[41] Trades are reset on a weekly basis, so that the models are always either 5% long or short. Clearly, this is a very simplistic trading rule and we could devise more far complicated alternatives. But keeping the rule relatively transparent allows for easier interpretation of performance results. We examine the performance of alternative rules below.

Performance data for this trading rule are reported in Table 13. These indicate that this strategy is profitable for all three exchange rates. IRs range from 0.37 for sterling-dollar to 0.74 for euro-dollar; sortino ratios-calculated using semi-standard deviation-are higher for each exchange rate, indicating that the distribution of returns is skewed to the downside. Directional success statistics-that calculate the number of weeks in which forecasts correctly predict the sign of exchange rate returns over the subsequent week, regardless of the magnitude of the change-indicate similar differences in forecasting accuracy between the models, with the euro-dollar model again most successful, predicting the sign of the subsequent week's return correctly 57% of the time. Less appealing is the comparison of annualised weekly returns with the 95% confidence interval: the fact that the 95% confidence interval encompasses zero raises a question as to the inherent skill of performance results based upon this naive trading rule.

An important characteristic of any trading strategy is the extent to which models for various exchange rates, when combined together within a portfolio,

[39] As determined by an informal survey of market participants. A notional 5% position size was selected for illustrative purposes only. We do not include interest rate carry in our simulations in order to better isolate the performance of underlying trading strategies.

[40] Close is defined as 4pm London time.

[41] We focus upon a 4-week forecast horizon as this traditionally offers the best liquidity for traders.

generate diversification benefits that improve the performance of this total portfolio. These benefits are usually measured in terms of the reduction of total risk of the trading strategy achieved through combining two or more models without any loss of return. The extent of diversification benefits to any trading strategy will be greater the lower is the correlation between return streams to component models; similarly, diversification benefits rise with the number of uncorrelated models within a portfolio.[42] From Table 13 the diversification benefits available within the Clarida-Taylor framework are clear: when our three exchange rate models are combined into a naive, equally weighted portfolio the IR of the combined strategy rises to 0.85 compared with 0.74 for the best performing individual model (euro-dollar). Diversification benefits are also apparent from the increase in sterling and sortino ratios, indicating that the expected payoff to this trading strategy improves markedly when individual models are combined together into a single portfolio. Similarly, annualised weekly returns are now approximately equal to the 95% confidence interval, and the directional success ratio of the total portfolio rises to 61%, compared with 57% for the euro-dollar (the best performing individual model). Although the directional success ratios of the yen-dollar (51%) and sterling-dollar (47%) models were relatively unattractive, the benefits of incorporating these two models into a portfolio with euro-dollar are demonstrated by all of these metrics, and reflect the relatively low correlation between excess returns to each of these models.

Overall, these are strong results: using a simple trading rule the Clarida-Taylor framework is able to add value to investment portfolios under realistic trading conditions. Indeed, these data compare favourably to performance data reported above for the representative FRB strategy of Rosenberg and Farka (2001), as well as by currency managers more generally–Baldridge, Meath and Myers (2000) report that overlay managers achieved an average annual IR of 0.55 over the sample period 1989 to 2000–and equity managers-Lyons (2001) reports that US equity managers employing a buy and hold strategy have achieved an annual IR of 0.40 over an (unspecified) fifty year period.

Table 15 reports IRs by calendar year for each of the three exchange rate models and the equally weighted portfolio. These data indicate that using the basic trading strategy the performance of the euro-dollar model has been consistent across the whole sample, with particularly strong performance in 1999 and 2002; these performance data suggest that the initial depreciation of the euro against the dollar, although not widely predicted by commentators (Consensus Economics, 1999) was actually consistent with the behaviour of fundamental variables. The performance of the yen-dollar and, particularly, sterling-dollar models has been more varied, and perhaps highlights the limits of the naive trading strategy we have chosen to adopt more than the Clarida-Taylor framework: nonetheless, this framework is likely to perform best in trending, rather than range-bound markets. For instance, in both 1999 and 2001 during which the sterling-dollar model loses money relative to the underlying benchmark re-

[42] The number of independent models, or "bets", indicates the breadth of a portfolio (Grinold and Kahn, 1999). Clearly, more breadth is desirable.

turn, the sterling-dollar exchange rate traded in a tight, ten figure range; over the course of 2002 and 2003 as a whole, when the model achieved annual IRs in excess of 1.0, sterling appreciated by forty big figures against the dollar.

There are a number of ways that the performance of the Clarida-Taylor framework may be improved within an investment portfolio context. These include adopting a more sophisticated trading rule that better extracts the information inherent in spot exchange rate forecasts, adopting best practice portfolio management techniques available to investors, and using information not captured by the Clarida-Taylor models in a qualitative framework to augment the information capture of these fundamental-based forecasting models.

In the previous chapter we discussed the behaviour of interbank dealers who manage foreign currency trading positions using a range of techniques, including tight stop-loss limits; these limits ensure that losing trades are cut at an early stage, while affording the trader confidence to run his profits.[43] Table 14 reports the performance of the previous naive buy-sell trading rule supplemented by an explicit intra-week stop-loss limit set at 10 basis points. Its impact upon IRs for euro-dollar, yen-dollar and the total portfolio is striking; the IR of the portfolio rises from 0.85 to 1.17; sterling and sortino ratios at the portfolio level increase markedly as well. In addition, the annualised weekly return from this strategy now exceeds the 95% confidence interval, suggesting inherent skill, and the longest period of underperformance has been reduced from 95 to 72 weeks. The benefits of an explicit loss-limit appear clear from these data: a combination of the Clarida-Taylor framework and prudent management of active currency hedges using tight stop-loss limits generates a substantial performance improvement upon any of the strategies discussed above.

The trading rule we have employed so far is relatively naive, taking a long (short) position in the domestic currency whenever the 4-week ahead forecast of the spot exchange rate exceeds (is below) the comparable forward rate. Positions are reset every week, with models either 5% long or short. We now consider an alternative strategy, whereby positions are in portfolios for four weeks, consistent with the horizon of the underlying exchange rate forecast, so that the absolute maximum position in portfolios at any time becomes 20%. In addition, we also introduce a no-trade threshold that eliminates trades from portfolios that are based upon tiny differences between the forecast level of spot and the 4-week forward rate; these differences are likely to be due to noise-perhaps caused by inaccuracies in data measurement-rather than evidence of profitable insight from the models.[44] The results of this augmented trading strategy are reported in Tables 16 and 17 (the latter also incorporating a 10 basis point stop-loss limit) and indicate that performance is markedly improved for the euro-dollar and yen-dollar models: IRs and Sortino ratios are higher; directional success rates also increase with this augmented strategy relative to the original, basic

[43] Clearly, there is a risk that the stop-loss limit may periodically cut a losing position just before the exchange rate moves in the traders' favours. On balance, the discipline provided by explicit stop-loss limits outweighs any negatives, particularly for risk-averse investors.

[44] We define noise as a differential between forecast spot and the 4-week forward rate of less than 0.0005 for euro-dollar and sterling-dollar, and 0.005 for yen-dollar.

trading rule; and for euro-dollar annualised returns exceed the 95% confidence interval. For sterling-dollar, by contrast, this strategy causes a deterioration in the profitability of the Clarida-Taylor approach. In addition, there is an absence of diversification benefits when all three models are combined into an equally weighted portfolio. These benefits return, however, when we augment this strategy to include an explicit intra-week stop-loss limit.[45] In this case, IRs, sortino and sterling ratios are all higher than for each of the individual models. These results are reported in Table 17. The IR of the portfolio rises from 1.71 to 2.53, with annualised weekly returns more than twice as large as the 95% confidence level, suggesting a high level of skill implicit within this strategy. Sterling ratios for the portfolio and individual exchange rate models are more attractive under this augmented trading strategy in the presence of a stop-loss limit, suggesting that expected payouts are much improved relative to the maximum drawdown experienced during the simulation period, and sortino ratios increase as well, reflecting the truncation of negative returns due to the stop-loss limit. In addition, the sterling-dollar model IR is positive once more, although this model's performance remains lacklustre compared with euro-dollar and yen-dollar models.

Further improvements in the profitability of Clarida-Taylor exchange rate forecasts may also be achieved by resort to best practice portfolio optimisation techniques that seek to achieve an optimal portfolio by selecting an efficient set of weights with which to combine our three exchange rate models, rather than equally weighting the three models. We examine three alternative optimisation techniques:[46] first, a traditional Markowitz approach (Markowitz, 1959), wherein optimal portfolio weights are chosen to minimise portfolio risk-as measured by standard deviation of returns-subject to achievement of some given level of expected returns, $E[R_t]$, where historical return is used as a proxy for $E[R_t]$; second, minimisation of portfolio risk without reference to historical return correlations; third, maximisation of $E[R_t]$ subject to the following constraint, C,

$$C = \beta\sigma + (1 - \beta)Drawdown \tag{55}$$

where the value of β is chosen to reflect the sensitivity of portfolio managers to risk and drawdown (JPMorgan Fleming Asset Management, 2003b).[47] The third approach explicitly recognises that risk-averse investors are sensitive to both the volatility and drawdown profile of cumulative returns and are not interested only in maximising expected returns subject to some given level of volatility. Incorporating a measure of drawdown into the portfolio optimisation problem therefore ensures that efficient portfolio weights are skewed towards models that generate relatively consistent returns relative to benchmark, and

[45] Stop loss limits attached to larger total position sizes may be problematic for larger investors-for instance, a major currrency overlay manager-to execute without impacting transaction costs (via a widening in spreads). Consequently, these results are more relevant to Hedge Fund or CTA investors for whom implied transaction size will remain appropriate.

[46] I thank Yazann Romahi for the use of his excel portfolio construction tool.

[47] We select a β of 0.5, for illustrative purposes.

that models susceptible to lumpy-positive and negative-excess returns are penalised; equivalently, this optimisation approach will favour component models with a relatively high sterling ratio.

We then additionally impose two, mutually exclusive portfolio constraints: first an annual return target of 5%; second, a maximum drawdown target of 2%. Once efficient portfolio weights have been determined in the first stage of the portfolio construction process under any of the three approaches described above, a linear scaling factor is then applied to these weights to achieve the secondary objective. This means that the performance of both portfolios in terms of all metrics except annualised return and maximum drawdown will be identical.

Efficient portfolio weights are reported in tables 18 and 19 (using the basic trading strategy) and 20 and 21 (augmented trading strategy). On the basis of IR statistics, the results indicate that only the Markowitz approach achieves a-small-improvement upon the performance of the equally weighted portfolio without stop-loss limits, on the basis of IRs, sterling and sortino ratios. But once stop-loss limits are incorporated into the equally weighted portfolio, its performance substantially exceeds the efficient-weight portfolios. Although the use of explicit intra-week stop-loss limits is not appropriate for portfolios that implement very large active currency hedges-because of liquidity management implications-our results have indicated the positive impact that they can exert upon portfolio performance where feasible to implement, consistent with the sensitivities of typical risk-averse investors; the corollary of an explicit stop-loss limit for larger portfolios is the drawdown-weighted efficient portfolio approach discussed above.

Finally, the quality of Clarida-Taylor forecasts may be improved by incorporating a qualitative assessment into the weekly trading decision to take account of non-fundamental exchange rate determinants not captured by this framework. These factors may include central bank foreign exchange intervention-a particularly topical example, given the extent of intervention against the yen during 2003 and into 2004 by the Bank of Japan-as well as strategic shifts in benchmark hedge ratios driven by change in portfolio risk-return preferences rather than any new fundamental data arrival: according to market anecdote this factor, which is extremely difficult to capture in a quantitative model, was an important determinant of the appreciation of euro-dollar during 2003. By taking a view on the incidence and relative importance of this type of factor, qualitative managers could be allowed to override the Clarida-Taylor models during infrequent-and generally short-time periods when fundamental factors may be of secondary importance to the sign and size of weekly exchange rate returns.

3.4 Conclusion

For two decades after publication of the seminal Meese-Rogoff (1983a, b) papers little evidence emerged to contradict their conclusion that fundamental-based exchange rate forecasting models are inferior to a random walk prediction (for a survey, see *Journal of International Economics*, 2003). As we discuss in the previous chapter, this conclusion also applies to much of the emerging literature on market microstructure, and particularly models that seek to exploit order flow data on a real-time basis. By contrast, investors have persistently demonstrated an ability to add value from foreign exchange trading, often from relatively naive trading rules based upon FRB and technical-based momentum strategies rather than sophisticated fundamental-based forecasting models (Baldridge, Meath and Myers, 2000; Hersey and Minnick, 2000). We suggest two reasons why the conclusions of academic and investor research may not be mutually exclusive: first, different focus in terms of forecasts (point versus directional); second, different performance metrics by which the quality of forecasts is assessed (MAFEs and RMSFEs versus profitability).

Recently, Clarida and Taylor (1997) and Clarida, Sarno, Taylor and Valente (2003) have reported the first serious challenge to the Meese-Rogoff evidence. Their models are based on a set of uncontroversial motivating assumptions and exploit information implicit in the forward rate term structure within a VECM system estimated by FIML. The models achieve improvements in forecast accuracy relative to a naive random walk of 50%-70% depending upon forecast horizon. In isolation this result means little in the context of investment portfolio management: improving upon the accuracy of random walk point forecasts does not guarantee persistent profits from foreign exchange trading. But using two simple trading rules, with associated portfolio construction tools, we demonstrate that the Clarida-Taylor framework can profitably be applied to investment portfolios, generating returns persistently in excess of an underlying strategic benchmark for euro-dollar, yen-dollar and sterling-dollar. To the best of the present authors' knowledge these results are the first demonstration of an ability to marry together academic and investor strands of exchange rate research under realistic transaction costs, position limits and publication lags within a profitable trading strategy. Furthermore, the simulated trading results that we report for the Clarida-Taylor framework appear superior to the performance of a traditional FRB strategy that is widely applied across the investor community, and also seem more consistent with investor sensitivities to the risk and drawdown characteristics of excess returns than is traditionally the case for FRB strategies.

Clarida, Sarno, Taylor and Valente (2003) extend the original results of Clarida and Taylor (1997) to a Markov-switching framework to exploit the inherent non-linearities present in the short-term, dynamic relationship between spot and forward exchange rates, and report an additional reduction in RMSFEs relative to a naive random walk model beyond that achieved by Clarida and Taylor (1997). These results suggest that the profitability of our trading rules may also be improved further by resort to this extension. This is a task that we leave to future research.

References

[1] Baillie, T.R. and T. Bollerslev (2000), The Forward Premium Anomaly is Not as Bad as You Think. Journal of International Money and Finance, 19, 471-488.

[2] Baldridge, J., B. Meath, and H. Myers (2000), "Capturing Alpha Through Active Currency Overlay", Frank Russell Company.

[3] Bansal, R. and M. Dahlquist (1999), The Forward Premium Puzzle: Different Tales from Developed and Emerging Economies. CEPR Working Paper #2169.

[4] Bank for International Settlements (2004), Triennial Central Bank Survey of Foreign Exchange and Derivatives Market Activity in April 2004. Preliminary Results. www.bis.org.

[5] Bank of England (1999), Economic Models at the Bank of England. Bank of England.

[6] Barkoulas, J.T., C.F. Baum and A. Chakraborty (2000), Forward Premiums and Market Efficiency: Panel Unit-Root Evidence from the Term Structure of Forward Premiums. Boston College Working Paper #461.

[7] Beveridge, S. and C.R. Nelson (1991), A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the "Business Cycle". Journal of Monetary Economics, 7, pp.151-74.

[8] Bilson, J.F.O. (1981), The "Speculative Efficiency" Hypothesis. Journal of Business, vol. 54, 3, 435-451.

[9] Carlson, J.A. & C.L. Osler (1999), Determinants of Currency Risk Premiums (sic.). Federal Reserve Bank of New York. http://www.newyorkfed.org/research/staff_reports/sr70.pdf.

[10] Chaboud, A. P. and J. H. Wright (2003), Uncovered Interest Parity: It Works, But Not for Long. International Finance Discussion Papers #752. Board of Governors of the Federal Reserve System.

[11] Clarida, R.H. and M.P. Taylor (1993), The Term Structure of Forward Exchange Rate Premiums and the Forecastability of Spot Exchange Rates: Correcting the Errors. NBER Working Paper #4442.

[12] Clarida, R.H. and M.P. Taylor (1997), The Term Structure of Forward Exchange Rate Premiums and the Forecastability of Spot Exchange Rates: Correcting the Errors. Review of Economics and Statistics, 79, 353-370.

[13] Clarida, R.H., L. Sarno, M.P. Taylor and G. Valente (2003), The Out-of-Sample Success of Term Structure Models as Exchange Rate Predictors: A Step Beyond. Journal of International Economics, 60, 61-83.

[14] Clarke, R. and M. Kritzman (2000), Currency Management: Concepts and Practices. Blackwell Publishing.

[15] CME (2004). Chicago Mercantile Exchange. www.cme.com.

[16] Consensus Economics (1999), Foreign Exchange Consensus Forecasts. http://www.consensuseconomics.com/.

[17] Crowder, W.J. (1994), Foreign Exchange Market Efficiency and Common Stochastic Trends. Journal of International Money and Finance, 13, 5, 551-564.

[18] Engel, C. and J.D. Hamilton (1990), Long Swings in the Dollar: Are They in the Data and Do Markets Know it? American Economic Review, 80, 689-713.

[19] Engel, C. (1996), The Forward Discount Anomaly and the Risk Premium: A Survey of Recent Evidence. Journal of Empirical Finance, 3, 123-192.

[20] Evans, M.D.D. and K.K. Lewis (1991), Peso Problems and Heterogeneous Trading: Evidence form Excess Returns in Foreign Exchange and Euromarkets. NBER Working Paper #4003.

[21] Evans, M.D.D. and K.K. Lewis (1992), Trends in Expected Returns in Currency and Bond Markets. NBER Working Paper #4116.

[22] Evans, M.D.D. and R.K. Lyons (2002), Order Flow and Exchange Rate Dynamics. Journal of Political Economy, February, 170-180.

[23] Fama, E. (1984), Forward and Spot Exchange Rates, Journal of Monetary Economics, 14, 319-338.

[24] Finn, M.G. (1986) Forecasting the Exchange Rate: A Monetary or Random Walk Phenomenon? Journal of International Money and Finance, 5, 181-93.

[25] Flood, R.P. and A.K. Rose (1994), Fixes: Of the Forward Discount Puzzle. NBER Working Paper #4928.

[26] Flood, R.P. and M. P. Taylor (1996), Exchange Rate Economics: What's Wrong with the Conventional Macro Approach?, in The Microstructure of Foreign Exchange Markets, eds. J A. Frankel, G. Galli and A. Giovannini, National Bureau of Economic Research. Chicago: University of Chicago Press.

[27] Froot, K.A. and J. A. Frankel (1986), Interpreting Tests of Forward Discount Bias Using Survey Data on Exchange Rate Expectations. NBER Working Paper #1963.

[28] Froot, K.A. and R.H. Thaler (1990), Anomalies: Foreign Exchange. Journal of Economic Perspectives, Vol. 4, 3, 179-192.

[29] Granger, C.W.J. (1983), Cointegrated Variables and Error Correcting Models. UCSD Discussion Paper.

[30] Granger, C.W.J. (1999), Aspects of Research Strategies for Time Series Analysis. Presentation to the Conference on New Developments in Time Series Economics, New Haven.

[31] Grinold, R.C. and R.N. Kahn (1999), Active Portfolio Management. MacGraw Hill.

[32] Hersey, B. and J. Minnick (2000), Currency Overlay. Global Pensions. February.

[33] Isard, P. (1995), Exchange Rate Economics. Cambridge Surveys of Economic Literature.

[34] JPMorgan Fleming Asset Management (2002), Currency Strategy Note. Unpublished Mimeo.

[35] JPMorgan Fleming Asset Management, (2003a), Currency Overlay. Unpublished Mimeo.

[36] JPMorgan Fleming Asset Management (2003b), Factor Risk Budgeting Using Drawdown. Unpublished Mimeo.

[37] JPMorgan Securities (1999), Introducing Our New Liquidity and Risk Premia Index. Global FX and Precious Metals Research, 25 August.

[38] Journal of International Economics (2003), Empirical Exchange Rate Models. Vol. 60, 1, 1-234.

[39] Kaminsky, (1993), Is There a Peso Problem? Evidence from the Dollar/Pound Exchange Rate, 1976-1987. American Economic Review, Vol. 83, 3, 450-72.

[40] Kritzman, M. (1993), The Optimal Hedging Policy with Biased Forward Rates. Journal of Portfolio Management.

[41] Laxton, D.M., P. Isard, H. Faruqee, E.S. Prasad and B. Turtleboom (1998), Multimod Mark III: The Core Dynamic and Steady State Model. IMF Occasional Paper #164. http://www.imf.org/external/pubs/ft/op/op164/index.htm

[42] Lewis, K.K. (1991), Was There a Peso Problem in the US Terms Structure of Interest Rates: 1979-1982? International Economic Review, Vol. 32, 159-173.

[43] Lewis, K.K. (1994), Puzzles in International Finance Markets. NBER Working Paper #4951.

[44] Lyons, R.K. (2001), The Microstructure Approach to Exchange Rates. Cambridge: MIT Press.

[45] Lyons, R.K. and A.K. Rose (1995), Explaining Forward Exchange Bias ... Intraday. Journal of Finance, 50, 1321-1329.

[46] MacDonald, R. and M.P. Taylor (1993) The Monetary Approach to the Exchange Rate: Rational Expectations, Long-Run Equilibrium and Forecasting. International Monetary Fund Staff Papers, 40, 89-107.

[47] Markowitz, H.M. (1959), Portfolio Selection: Efficient Diversification of Investment. Cowles Foundation Monograph 16. New Haven, Conn.: Yale University Press.

[48] Marston, (1994), Tests of Three Parity Conditions: Distinguishing Risk Premia and Systematic Forecast Errors. NBER Working Paper #4923.

[49] Maynard and Phillips (2001), Rethinking an old empirical puzzle: Econometric evidence on the forward discount anomaly. Journal of Applied Econometrics, Vol. 16, 6, 671-708.

[50] Meese, R.A. and K. Rogoff (1983a), Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample? Journal of International Economics, 14, 3-24.

[51] Meese, R.A. and K. Rogoff (1983b), The Out-of-Sample Failure of Empirical Exchange Rate Models: Sampling Error or Misspecification? in Frenkel, J.A. (ed.), Exchange Rate and International Economics, University of Chicago Press.

[52] Meredith and Ma (2002), The Forward Premium Puzzle Revisited. IMF Working Paper #02/28. http://www.imf.org/external/pubs/cat/longres.cfm?sk=15624.0.

[53] Mussa, M. (1984), The Theory of Exchange Rate Determination, in Bilson, J. and R. Marston (eds.), Exchange Rate Theory and Practice, National Bureau of Economic Research Conference Report, Chicago and London: University of Chicago Press, 13-58.

[54] Rogoff, K. (1979), Expectations and Exchange Rate Volatility. Unpublished Ph.D. Thesis. Massachusetts Institute of Technology.

[55] Rosenberg, M. and E. Farka (2001), Forward Rate Bias Strategy. Deutsche Bank.

[56] Sarno, L. and M. P. Taylor (2002), The Economics of Exchange Rates. Cambridge: Cambridge University Press.

[57] Strange, B. (1998), Currency Managers Show Consistency. Pension and Investments. June.

[58] Taylor, M.P. (1987), Covered Interest Parity: A High Frequency High Quality Data Study. Economica 54, 429-38.

[59] Taylor, M.P. (1989), Covered Interest Arbitrage and Market Turbulence. Economic Journal 99, 376-91.

[60] Taylor, M.P. (1995),The Economics of Exchange Rates. Journal of Economic Literature 13, 33-47.

[61] Taylor, M.P., D. Peel and L. Sarno (2001), Non-Linear Mean Reversion in Real Exchange Rates: Towards a Solution to the Purchasing Power Parity Puzzles. International Economic Review, 42, 1015-42.

[62] Tong, H. (1990), Nonlinear Time Series: A Dynamic System Approach. Clarendon Press.

[63] UBS (2004), Risk Aversion Index. UBS.

Table 1: Uncovered Interest Parity Regressions

Mark		*Yen*		*Sterling*	
α	-0.0017	α	-0.0017	α	0.0041
	(0.0019)		(0.0004)		(0.0017)
β	-0.0843	β	-0.0405	β	-0.1793
	(0.0592)		(0.0107)		(0.0760)
$Adj.R^2$	0.0063	$Adj.R^2$	0.0282	$Adj.R^2$	0.0235
$Akaike$	-4.0545	$Akaike$	-7.1862	$Akaike$	-4.2066
$Normality^*$	0.0045	$Normality^*$	0.0000	$Normality^*$	0.0000
$Breusch^*$	0.0000	$Breusch^*$	0.0000	$Breusch^*$	0.0000
$ARCH(4)^*$	0.0000	$ARCH(4)^*$	0.0000	$ARCH(4)^*$	0.0000

Notes: We test the validity of UIP for mark-dollar, yen-dollar and sterling-dollar by estimating equation (41) above, where $k = 4$. Equations estimated over the sample January 1979-December 2003. Estimated standard errors incorporate Newey-West consistent standard errors. Normality is the Jarque-Bera test for residual normality. Breusch is the Breusch-Godfrey test for residual serial correlation, ARCH(4) is a 4-lag test for residual heteroskedasticity. * signifies p-value.

Table 2: Augmented Dickey Fuller (ADF) Unit Root Tests

		I(0)			
		Constant	*Lag Length*	*Constant and Trend*	*Lag Length*
Mark	S	−1.1154	0	−1.8504	0
	$LF1$	−1.1252	0	−1.8662	0
	$LF3$	−1.1453	0	−1.8951	0
	$LF6$	−1.1860	0	−1.9514	0
	$LF12$	−1.3653	4	−2.0461	4
Yen	S	−0.7965	2	−1.9441	2
	$LF1$	−0.7851	2	−1.9622	2
	$LF3$	−0.7646	2	−2.0543	4
	$LF6$	−0.8533	4	−2.0908	4
	$LF12$	−0.6645	1	−2.1460	4
Sterling	S	−1.8928	0	−1.7598	0
	$LF1$	−1.8922	0	−1.7552	0
	$LF3$	−1.8934	0	−1.7517	0
	$LF6$	−1.8949	0	−1.7458	0
	$LF12$	−1.8987	0	−1.7389	0

		I(1)			
		Constant	*Lag Length*	*Constant and Trend*	*Lag Length*
Mark	S	−31.8243	0	−31.8162	0
	$LF1$	−31.9171	0	−31.9093	0
	$LF3$	−32.1374	0	−32.1303	0
	$LF6$	−14.7086	3	−14.7083	3
	$LF12$	−14.7662	3	−14.7663	3
Yen	S	−20.6575	1	−20.6484	1
	$LF1$	−20.7383	1	−20.7295	1
	$LF3$	−21.016	1	−21.0077	1
	$LF6$	−14.4635	3	−14.4576	3
	$LF12$	−36.6415	0	−36.6287	0
Sterling	S	−31.3590	0	−31.3562	0
	$LF1$	−31.3907	0	−31.3880	0
	$LF3$	−31.4975	0	−31.4950	0
	$LF6$	−31.6032	0	−31.6012	0
	$LF12$	−31.7323	0	−31.7311	0

Table 3: Johansen Cointegration Tests

Mark

Linear Deterministic Trend; Lag Length: 9

No. Hypothesised CEs	*Eigenvalue*	*Trace Statistic*	5% *Critical Value*	1% *Critical Value*
*None***	0.1092	229.7711	68.52	76.07
*At Most One***	0.0692	110.2309	47.21	54.46
*At Most Two***	0.0247	36.1070	29.68	35.65
At Most Three	0.0080	10.1996	15.41	20.04
At Most Four	0.0017	1.7999	3.76	6.65

Yen

Linear Deterministic Trend; Lag Length: 6

No. Hypothesised CEs	*Eigenvalue*	*Trace Statistic*	5% *Critical Value*	1% *Critical Value*
*None***	0.1192	259.0769	68.52	76.07
*At Most One***	0.0799	127.6388	47.21	54.46
*At Most Two***	0.0227	41.4172	29.68	35.65
*At Most Three***	0.0140	17.6356	15.41	20.04
At Most Four	0.0028	2.9724	3.76	6.65

Sterling

No Deterministic Trend; Lag Length: 5

No. Hypothesised CEs	*Eigenvalue*	*Trace Statistic*	5% *Critical Value*	1% *Critical Value*
*None***	0.1421	295.8124	59.46	66.52
*At Most One***	0.0801	136.8654	39.89	45.58
*At Most Two***	0.0341	50.2618	24.31	29.75
*At Most Three**	0.0124	14.1855	12.53	16.31
At Most Four	0.0011	1.1963	3.84	6.51

Table 4: Test of Overidentifying Restrictions on Estimated Cointegrating Coefficients

	$\chi^2\,(g)$	$P-value$
Mark	11.9081	$1.8x10^{-2}$
Yen	20.5044	$3.0x10^{-4}$
Sterling	30.5640	$4.0x10^{-6}$

Notes: The test is a χ^2 version of the overidentifying restriction on the $\beta^{'}$ matrix described in equation (54) above. g is the number of restrictions imposed, which in this case is four for all exchange rates. The test is conditional on the presence of four linearly independent cointegrating vectors.

Table 5: Long-Run Estimated Cointegrating Coefficients

	Coefficient	Standard Error
Mark	−1.0986	0.0448
Yen	−1.0095	0.0168
Sterling	−1.0326	0.0119

Notes: The table reports the estimated cointegrating coefficient for the 12-month forward rate for each exchange rate. Other parameters restricted to equal unity (spot) and zero (other forward rates).

Table 6: Optimised VECM System for Mark-Dollar

	Dependent Variable:				
	s	f^4	f^{13}	f^{26}	f^{52}
c	0.0003	--	--	--	0.0001
	(0.0001)				$(5.92E-05)$
Δs_{t-1}	--	--	--	−0.0082	--
			(0.0194)	(0.0027)	
Δf^4_{t-1}	--	--	--	−0.1390	−0.2834
			(0.0264)	(0.0239)	(0.0781)
Δf^{13}_{t-1}	1.2853	1.2047	1.2377	1.5579	1.6874
	(0.4255)	(0.3274)	(0.3158)	(0.3076)	(0.3385)
Δf^{26}_{t-1}	−1.5925	−1.4111	−1.3624	−1.4098	−1.0426
	(0.5407)	(0.3478)	(0.3223)	(0.3045)	(0.3309)
Δf^{52}_{t-1}	0.3329	0.2109	0.1271	--	−0.3638
	(0.1670)	(0.0621)	(0.0379)		(0.0525)
$(s-f^4)_{t-1}$	--	1.6394	1.3403	1.3028	0.9258
		(0.0433)	(0.0444)	(0.0516)	(0.0762)
$(s-f^{13})_{t-1}$	0.0620	−0.5407	--	--	0.7333
	(0.0320)	(0.0265)	(0.0460)		(0.0750)
$(s-f^{26})_{t-1}$	--	--	−0.2752	−0.2098	−0.6745
			(0.0162)	(0.0135)	(0.0585)
$(s-f^{52})_{t-1}$	--	--	0.0286	--	0.0894
			(0.0033)		(0.0145)
$Adj.R^2$	0.9947	0.9952	0.9951	0.9949	0.9943
SSR	0.2405	0.2405	0.2130	0.2142	0.2250

Table 6 (cont.): Optimised VECM System for Yen-Dollar

	Dependent Variable:				
	s	f^4	f^{13}	f^{26}	f^{52}
c	−0.0033 (0.0008)	−0.0034 (0.0008)	−0.0035 (0.0009)	−0.0037 (0.0009)	−0.0045 (0.0010)
Δs_{t-1}	−−	−−	0.1720 (0.0194)	0.2779 (0.0224)	0.3909 (0.0345)
Δf^4_{t-1}	−−	0.1589 (0.0430)	0.1175 (0.0264)	0.0858 (0.0193)	−−
Δf^{13}_{t-1}	−−	−0.1531 (0.0440)	−0.2885 (0.0274)	−0.1449 (0.0414)	−0.2818 (0.0695)
Δf^{26}_{t-1}	−−	−−	−−	−0.2182 (0.0222)	0.2233 (0.0672)
Δf^{52}_{t-1}	−−	−−	−− (0.2370)	−−	−0.3353 (0.0208)
$(s-f^4)_{t-1}$	−−	1.1091 (0.0569)	0.1662 (0.0337)	0.1321 (0.0204)	−−
$(s-f^{13})_{t-1}$	−0.7838 (0.0803)	−0.5485 (0.0692)	0.4902 (0.0460)	−−	−−
$(s-f^{26})_{t-1}$	−0.5554 (0.0642)	−−	−0.4249 (0.0362)	−−	−0.0973 (0.0395)
$(s-f^{52})_{t-1}$	−−	0.1260 (0.0250)	0.1676 (0.0203)	0.0864 (0.0221)	0.1699 (0.0387)
$Adj.R^2$	0.9976	0.9977	0.9976	0.9975	0.9973
SSR	0.2585	0.2438	0.2517	0.2571	0.2785

Table 6 (cont.): Optimised VECM System for Sterling-Dollar

	Dependent Variable:				
	s	f^4	f^{13}	f^{26}	f^{52}
c	-0.0024	-0.0025	-0.0025	-0.0025	-0.0026
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Δs_{t-1}	-1.4019	1.4009	-1.3878	-1.3991	-1.4185
	(0.1557)	(0.1633)	(0.1635)	(0.1642)	(0.1652)
Δf^4_{t-1}	-0.6451	-0.1854	--	--	--
	(0.1150)	(0.0223)			
Δf^{13}_{t-1}	3.6765	3.0722	2.6872	2.6990	2.4404
	(0.3869)	(0.3213)	(0.3168)	(0.3094)	(0.3040)
Δf^{26}_{t-1}	-0.6306	-0.5060	-0.3138	-0.3566	--
	(0.2799)	(0.1819)	(0.1465)	(0.0901)	
Δf^{52}_{t-1}	-1.0001	-0.9800	-0.9884	-0.9480	-1.0283
	(0.2514)	(0.2420)	(0.2370)	(0.2334)	(0.2327)
$(s-f^4)_{t-1}$	--	1.4582	1.1812	1.2996	1.2101
		(0.1227)	(0.1168)	(0.1281)	(0.1602)
$(s-f^{13})_{t-1}$	--	7.7409	8.1011	7.7322	8.0407
		(0.2392)	(0.2427)	(0.2647)	(0.3211)
$(s-f^{26})_{t-1}$	-7.7435	-8.0265	-8.1887	-7.8843	-8.1830
	(0.2396)	(0.2933)	(0.2968)	(0.3080)	(0.3378)
$(s-f^{52})_{t-1}$	2.0466	2.1332	2.1488	2.0775	2.1679
	(0.1386)	(0.1434)	(0.1436)	(0.1445)	(0.1481)
$Adj.R^2$	0.9890	0.9892	0.9893	0.9893	0.9893
SSR	0.2491	0.2450	0.2447	0.2452	0.2461

Table 7: MAFE and RMSFE Analysis for Mark-Dollar

Horizon (weeks)	*MAFE*	*RMSFE*	*Horizon (weeks)*	*MAFE*	*RMSFE*
1	0.6805	0.6948	27	0.3374	0.3160
2	0.5433	0.5579	28	0.3414	0.3114
3	0.4719	0.4992	29	0.3431	0.3065
4	0.4600	0.4849	30	0.3451	0.3026
5	0.4033	0.4493	31	0.3510	0.2996
6	0.3695	0.4173	32	0.3476	0.2942
7	0.34664	0.3932	33	0.3419	0.2888
8	0.3292	0.3751	34	0.3331	0.2838
9	0.3172	0.3620	35	0.3244	0.2789
10	0.3046	0.3482	36	0.3244	0.2740
11	0.2897	0.3381	37	0.3179	0.2688
12	0.2809	0.3279	38	0.3108	0.2627
13	0.2769	0.3214	39	0.3007	0.2572
14	0.2743	03165	40	0.2943	0.2512
15	0.2719	0.3138	41	0.2866	0.2451
16	0.2725	0.3128	42	0.2782	0.2385
17	0.2727	0.3136	43	0.2706	0.2310
18	0.2789	0.3192	44	0.2621	0.2240
19	0.2813	0.3210	45	0.2578	0.2200
20	0.2868	0.3245	46	0.2549	0.2169
21	0.2931	0.3281	47	0.2526	0.2142
22	0.2976	0.3282	48	0.2500	0.2120
23	0.3032	0.3266	49	0.2512	0.2105
24	0.3080	0.3232	50	0.2525	0.2102
25	0.3167	0.3223	51	0.2528	0.2092
26	0.3254	0.3192	52	0.2533	0.2080

Notes: MAFE is defined as the Mean Absolute Forecast Error; RMSFE is defined as the Root Mean Squared Forecast Error. The table reports the ratio of MAFEs and RMSFEs for n-period ahead exchange rate forecasts from the Clarida-Taylor framework relative to a naive random walk. Consequently, a lower number implies that Clarida-Taylor forecast accuracy is rising relative to the random walk.

Table 8: Diebold-Mariano Test Statistics for Mark-Dollar

Horizon (*weeks*)	*P-value*	*Horizon* (*weeks*)	*P-value*
1	1.9×10^{-4}	27	1.1×10^{-3}
2	5.0×10^{-8}	28	4.2×10^{-3}
3	5.0×10^{-9}	29	9.5×10^{-3}
4	1.3×10^{-6}	30	1.5×10^{-2}
5	3.5×10^{-7}	31	$2..2 \times 10^{-2}$
6	1.0×10^{-7}	32	2.4×10^{-2}
7	9.5×10^{-8}	33	2.4×10^{-2}
8	7.5×10^{-8}	34	2.3×10^{-2}
9	1.1×10^{-7}	35	2.1×10^{-2}
10	5.2×10^{-7}	36	3.9×10^{-2}
11	1.0×10^{-6}	37	1.7×10^{-2}
12	6.2×10^{-7}	38	1.5×10^{-2}
13	2.0×10^{-8}	39	1.2×10^{-2}
14	5.0×10^{-9}	40	1.0×10^{-2}
15	5.0×10^{-9}	41	8.2×10^{-3}
16	5.0×10^{-9}	42	6.6×10^{-3}
17	0	43	5.8×10^{-3}
18	0	44	5.4×10^{-3}
19	0	45	5.5×10^{-3}
20	0	46	5.8×10^{-3}
21	0	47	7.0×10^{-3}
22	0	48	7.8×10^{-3}
23	0	49	9.1×10^{-3}
24	5.5×10^{-8}	50	1.0×10^{-2}
25	4.0×10^{-6}	51	1.1×10^{-2}
26	1.1×10^{-4}	52	1.2×10^{-2}

Notes: The Diebold Mariano (DM) test of forecast accuracy indicates the significance of differences in forecast accuracy. The table reports associated p-values. A DM test p-value below 0.05 indicates that Clarida-Taylor exchange rate forecasts are more accurate forecasts than random walk forecasts at traditional significance levels. P-Values smaller than $1.0e - 10^{-10}$ reported as zero.

Table 9: MAFE and RMSFE Analysis for Yen-Dollar

Horizon (weeks)	*MAFE*	*RMSFE*	*Horizon (weeks)*	*MAFE*	*RMSFE*
1	0.6943	0.7109	27	0.2844	0.2676
2	0.5553	0.5721	28	0.2820	0.2668
3	0.4827	0.5077	29	0.2808	0.2654
4	0.4675	0.4945	30	0.2796	0.2651
5	0.4402	0.4650	31	0.2799	0.2670
6	0.4053	0.4313	32	0.2785	0.2675
7	0.3913	0.4085	33	0.2756	0.2678
8	0.3718	0.3849	34	0.2745	0.2692
9	0.3584	0.3691	35	0.2714	0.2708
10	0.3511	0.3570	36	0.2710	0.2719
11	0.3428	0.3431	37	0.2663	0.2712
12	0.3304	0.3282	38	0.2611	0.2686
13	0.3216	0.3184	39	0.2582	0.2675
14	0.3115	0.3084	40	0.2555	0.2657
15	0.2994	0.3006	41	0.2515	0.2631
16	0.2915	0.2915	42	0.2460	0.2606
17	0.2863	0.2863	43	0.2404	0.2563
18	0.2836	0.2833	44	0.2366	0.2548
19	0.2797	0.2787	45	0.2341	0.2559
20	0.2781	0.2771	46	0.2324	0.2586
21	0.2792	0.2778	47	0.2313	0.2608
22	0.2801	0.2758	48	0.2304	0.2616
23	0.2818	0.2732	49	0.2309	0.2641
24	0.2841	0.2714	50	0.2321	0.2676
25	0.2877	0.2710	51	0.2321	0.2704
26	0.2877	0.2695	52	0.2324	0.2717

Notes: MAFE is defined as the Mean Absolute Forecast Error; RMSFE is defined as the Root Mean Squared Forecast Error. The table reports the ratio of MAFEs and RMSFEs for n-period ahead exchange rate forecasts from the Clarida-Taylor framework relative to a naive random walk. Consequently, a lower number implies that Clarida-Taylor forecast accuracy is rising relative to the random walk.

Table 10: Diebold-Mariano Test Statistics for Yen-Dollar

Horizon (weeks)	*P-value*	*Horizon (weeks)*	*P-value*
1	5.8×10^{-4}	27	4.5×10^{-4}
2	1.7×10^{-6}	28	1.5×10^{-4}
3	1.2×10^{-7}	29	5.4×10^{-5}
4	1.8×10^{-5}	30	1.1×10^{-5}
5	1.5×10^{-4}	31	2.1×10^{-6}
6	3.6×10^{-4}	32	2.0×10^{-7}
7	8.0×10^{-4}	33	0
8	1.5×10^{-3}	34	6.6×10^{-4}
9	4.2×10^{-3}	35	4.6×10^{-4}
10	9.5×10^{-3}	36	3.0×10^{-4}
11	1.2×10^{-2}	37	1.6×10^{-4}
12	1.2×10^{-2}	38	8.7×10^{-5}
13	1.3×10^{-2}	39	4.1×10^{-5}
14	1.2×10^{-2}	40	1.6×10^{-5}
15	1.1×10^{-2}	41	5.3×10^{-6}
16	9.7×10^{-3}	42	1.4×10^{-6}
17	7.8×10^{-3}	43	3.2×10^{-7}
18	6.2×10^{-3}	44	7.0×10^{-8}
19	5.0×10^{-3}	45	1.5×10^{-8}
20	4.6×10^{-3}	46	0
21	4.4×10^{-3}	47	0
22	3.9×10^{-3}	48	0
23	3.2×10^{-3}	49	0
24	2.6×10^{-3}	50	0
25	1.8×10^{-3}	51	0
26	1.1×10^{-3}	52	0

Notes: The Diebold Mariano (DM) test of forecast accuracy indicates the significance of differences in forecast accuracy. The table reports associated p-values. A DM test p-value below 0.05 indicates that Clarida-Taylor exchange rate forecasts are more accurate forecasts than random walk forecasts at traditional significance levels. P-Values smaller than $1.0e - 10^{-10}$ reported as zero.

Table 11: MAFE and RMSFE Analysis for Sterling-dollar

Horizon (weeks)	*MAFE*	*RMSFE*	*Horizon (weeks)*	*MAFE*	*RMSFE*
1	1.0035	1.0095	27	0.5328	0.4785
2	0.7625	0.7750	28	0.5353	0.4658
3	0.7013	0.7024	29	0.5258	0.4528
4	0.6634	0.6760	30	0.5239	0.4414
5	0.6583	0.6644	31	0.5227	0.4335
6	0.6311	0.6487	32	0.5163	0.4235
7	0.6205	0.6442	33	0.5179	0.4164
8	0.5970	0.6334	34	0.5291	0.4143
9	0.5766	0.6276	35	0.5353	0.4135
10	0.5668	0.6172	36	0.5375	0.4070
11	0.5531	0.6098	37	0.5353	0.4070
12	0.5412	0.5959	38	0.5204	0.3992
13	0.5302	0.5774	39	0.5094	0.3910
14	0.5222	0.5656	40	0.4953	0.3840
15	0.5172	0.5555	41	0.4861	0.3759
16	0.5124	0.5514	42	0.4800	0.3685
17	0.5080	0.5439	43	0.4684	0.3591
18	0.5198	0.5423	44	0.4544	0.3512
19	0.5183	0.5365	45	0.4443	0.3456
20	0.5161	0.5319	46	0.4377	0.3413
21	0.5160	0.5302	47	0.4346	0.3381
22	0.5126	0.5254	48	0.4314	0.3357
23	0.5169	0.5194	49	0.4304	0.3355
24	0.5154	0.5059	50	0.4312	0.3347
25	0.5159	0.4979	51	0.4265	0.3345
26	0.5251	0.4915	52	0.4291	0.3329

Notes: MAFE is defined as the Mean Absolute Forecast Error; RMSFE is defined as the Root Mean Squared Forecast Error. The table reports the ratio of MAFEs and RMSFEs for n-period ahead exchange rate forecasts from the Clarida-Taylor framework relative to a naive random walk. Consequently, a lower number implies that Clarida-Taylor forecast accuracy is rising relative to the random walk.

Table 12: Diebold-Mariano Test Statistics for Sterling-Dollar

Horizon (weeks)	*P-value*	*Horizon (weeks)*	*P-value*
1	5.0×10^{-3}	27	0
2	0	28	0
3	4.9×10^{-7}	29	0
4	5.3×10^{-7}	30	0
5	3.2×10^{-7}	31	0
6	3.0×10^{-7}	32	5.3×10^{-7}
7	6.0×10^{-8}	33	1.7×10^{-6}
8	0	34	1.7×10^{-6}
9	0	35	1.4×10^{-6}
10	0	36	2.9×10^{-7}
11	0	37	3.6×10^{-7}
12	0	38	3.5×10^{-8}
13	3.0×10^{-4}	39	0
14	0	40	0
15	0	41	0
16	5.0×10^{-9}	42	5.0×10^{-9}
17	1.9×10^{-6}	43	0
18	2.3×10^{-6}	44	0
19	3.1×10^{-6}	45	3.0×10^{-9}
20	2.1×10^{-6}	46	3.7×10^{-7}
21	1.4×10^{-6}	47	7.3×10^{-7}
22	4.2×10^{-6}	48	9.2×10^{-7}
23	3.0×10^{-8}	49	2.0×10^{-7}
24	0	50	7.5×10^{-8}
25	0	51	4.5×10^{-8}
26	0	52	1.5×10^{-8}

Notes: The Diebold Mariano (DM) test of forecast accuracy indicates the significance of differences in forecast accuracy. The table reports associated p-values. A DM test p-value below 0.05 indicates that Clarida-Taylor exchange rate forecasts are more accurate forecasts than random walk forecasts at traditional significance levels. P-Values smaller than $1.0e - 10^{-10}$ reported as zero.

Table 13: Performance Data for Basic Trading Strategy - Individual Exchange Rate Models

	Euro	Yen	Sterling	Portfolio
Directional Success	57.44%	51.35%	47.71%	61.28%
Annualised Return	0.39%	0.26%	0.14%	0.79%
Annualised Risk	0.52%	0.51%	0.39%	0.93%
Information Ratio	0.74	0.51	0.37	0.85
95% Confidence	0.45%	0.45%	0.34%	0.81%
Maximum Drawdown	0.68%	1.09%	0.56%	1.03%
Long. U/perf. (weeks)	41	120	96	95
Sterling Ratio	0.56	0.24	0.26	0.76
Sortino Ratio	1.03	0.79	0.53	1.20

Notes: performance data assume an underlying notional +/-5% currency position in each of the three exchange rate models. Positions are generated by a naive buy-sell trading rule that is long (short) currency i versus the dollar whenever the level of the forecast 4-week spot exchange rate is above (below) the 4-week forward rate. Risk is the standard deviation of weekly returns. Information Ratio is the ratio of annualised returns to annualised risk. Maximum drawdown is the largest peak-to-trough decline in cumulative returns during the simulation period. Longest Underperformance is the longest number of weeks taken to recover the previous local peak in cumulative returns following an initial drawdown. Transaction costs assumed at 5 basis points round trip, on the basis of an informal survey of market practitioners.

Table 14: Performance Data for Basic Trading Strategy Incorporating Weekly Stop–Loss Limit within Individual Exchange Rate Models

	Euro	Yen	Sterling	Portfolio
Directional Success	56.32%	51.55%	47.51%	
Annualised Return	0.51%	0.35%	0.17%	1.03%
Annualised Risk	0.48%	0.49%	0.38%	0.89%
Information Ratio	1.05	0.72	0.44	1.17
95% Confidence	0.42%	0.43%	0.33%	0.77%
Maximum Drawdown	0.50%	0.90%	0.50%	0.82%
Long. U/perf. (weeks)	33	96	88	72
Sterling Ratio	1.02	0.39	0.34	1.25
Sortino Ratio	1.51	1.08	0.64	1.69

Notes: performance data assume an underlying notional +/-5% currency position in each of the three exchange rate models. Positions are generated by a naive buy-sell trading rule that is long (short) currency i versus the dollar whenever the level of the forecast 4-week spot exchange rate is above (below) the 4-week forward rate. Risk is the standard deviation of weekly returns. Information Ratio is the ratio of annualised returns to annualised risk. Maximum drawdown is the largest peak-to-trough decline in cumulative returns during the simulation period. Longest Underperformance is the longest number of weeks taken to recover the previous local peak in cumulative returns following an initial drawdown. Transaction costs assumed at 5 basis points round trip, on the basis of an informal survey of market practitioners.

Table 15: Information Ratios by Calendar Year for Basic and Stop-Loss Trading Strategy - Individual Exchange Rate Models

	Euro		Yen		Sterling		Portfolio	
	Basic	Stop Loss	Basic	Stop Loss	Basic	Stop Loss	Basic	Stop Loss
1999	1.03	1.06	−0.10	0.30	−0.31	−0.31	0.12	0.46
2000	0.38	0.74	0.72	0.81	0.07	0.14	0.61	0.97
2001	0.29	0.64	0.04	0.14	−0.11	0.11	−0.34	−0.06
2002	1.54	1.80	0.79	1.10	1.26	1.26	1.11	1.42
2003	0.83	1.34	1.33	1.41	1.04	1.04	1.59	1.90

Notes: performance data assume an underlying notional +/-5% currency position in each of the three exchange rate models. Positions are generated by a naive buy-sell trading rule that is long (short) currency i versus the dollar whenever the level of the forecast 4-week spot exchange rate is above (below) the 4-week forward rate. Risk is the standard deviation of weekly returns. Information Ratio is the ratio of annualised returns to annualised risk. Maximum drawdown is the largest peak-to-trough decline in cumulative returns during the simulation period. Longest Underperformance is the longest number of weeks taken to recover the previous local peak in cumulative returns following an initial drawdown. Transaction costs assumed at 5 basis points round trip, on the basis of an informal survey of market practitioners.

Table 16: Performance Data for Augmented Trading Strategy

	Euro	Yen	Sterling	Portfolio
Directional Success	58.24%	54.22%	45.59%	
Annualised Return	1.10%	0.72%	−0.12%	1.71%
Annualised Risk	1.01%	0.98%	0.76%	1.93%
Information Ratio	1.09	0.74	−0.15	0.88
95% Confidence	0.93%	1.01%	0.67%	1.68%
Maximum Drawdown	2.18%	3.52%	1.78%	5.36%
Long. U/perf. (weeks)	107	195	250	168
Sterling Ratio	0.50	0.21	−0.07	0.32
Sortino Ratio	1.54	1.06	−0.17	1.34

Notes: performance data assume an underlying notional +/-5% currency position is generated each week in each of the three exchange rate models. The trading rule is based upon a naive buy-sell trading rule that is long (short) currency i versus the dollar whenever the level of the forecast 4-week spot exchange rate is above (below) the 4-week forward rate. Positions are then held for four weeks, meaning that the maximum active hedge in portfolios is +/-20%. Risk is the standard deviation of weekly returns. Information Ratio is the ratio of annualised returns to annualised risk. Maximum drawdown is the largest peak-to-trough decline in cumulative returns during the simulation period. Longest Underperformance is the longest number of weeks taken to recover the previous local peak in cumulative returns following an initial drawdown. Transaction costs assumed at 5 basis points round trip, on the basis of an informal survey of market practitioners.

Table 17: Performance Data for Augmented Trading Strategy Incorporating Weekly Stop–Loss Limit within Individual Exchange Rate Models

	Euro	Yen	Sterling	Portfolio
Directional Success	58.24%	54.22%	47.51%	
Annualised Return	1.94%	1.13%	0.37%	7.21%
Annualised Risk	0.83%	0.89%	0.66%	2.85%
Information Ratio	2.34	1.27	0.57	2.53
95% Confidence	0.73%	0.79%	0.58%	2.48%
Maximum Drawdown	0.89%	2.36%	0.95%	3.13%
Long. U/perf. (weeks)	32	92	83	49
Sterling Ratio	2.18	0.48	0.40	2.30
Sortino Ratio	3.76	1.93	0.93	4.36

Notes: performance data assume an underlying notional +/-5% currency position is generated each week in each of the three exchange rate models. The trading rule is based upon a naive buy-sell trading rule that is long (short) currency i versus the dollar whenever the level of the forecast 4-week spot exchange rate is above (below) the 4-week forward rate. Positions are then held for four weeks, meaning that the maximum active hedge in portfolios is +/-20%. Risk is the standard deviation of weekly returns. Information Ratio is the ratio of annualised returns to annualised risk. Maximum drawdown is the largest peak-to-trough decline in cumulative returns during the simulation period. Longest Underperformance is the longest number of weeks taken to recover the previous local peak in cumulative returns following an initial drawdown. Transaction costs assumed at 5 basis points round trip, on the basis of an informal survey of market practitioners.

Table 18: Performance Data for Basic Trading Strategy - Efficient Portfolios with Annual Return Target

	Portfolio I	Portfolio II	Portfolio III
Annualised Return	5.00%	5.00%	5.00%
Annualised Risk	5.55%	6.91%	6.01%
Information Ratio	0.90	0.72	0.83
95% Confidence Interval	4.86%	6.05%	5.26%
Maximum Drawdown	6.38%	9.35%	5.80%
Longest Underperformance	38	78	44
Sterling Ratio	0.78	0.54	0.86
Sortino Ratio	1.24	1.04	1.18

Notes: Performance data assume an underlying 5% annualised return target, generated by a naive buy-sell trading rule that is long (short) currency i versus the dollar whenever the forecast 4-week spot exchange rate is above (below) the 4-week forward rate. Weights for each portfolio indicate the implied weekly position in each exchange rate and, by aggregation, weekly total portfolio positions relative to a neutral (strategic) currency benchmark. Portfolio I weights the three exchange rate models together on the basis of return and risk correlations (that is, within standard Markowitz mean-variance optimisation framework). Portfolio II combines the models on the basis of risk minimisation. Portfolio III weights models on the basis of risk and drawdown characteristics of simulated returns, and using the constraint in equation (56) above, with β set at 0.5. All portfolios constructed using a Visual Basic optimisation tool. Transaction costs assumed to be 5 basis points round trip, based upon an informal survey of market participants.

Portfolio I Weights: 9.67 Euro; 3.02 Yen; 2.81 Sterling.
Portfolio II Weights: 4.70 Euro; 7.86 Yen; 11.49 Sterling.
Portfolio III Weights: 6.95 Euro; 5.22 Yen; 8.09 Sterling.

Table 19: Performance Data for Basic Trading Strategy - Efficient Portfolios with Annual Drawdown Target

	Portfolio I	Portfolio II	Portfolio III
Annualised Return	1.57%	1.07%	1.75%
Annualised Risk	1.74%	1.48%	2.11%
Information Ratio	0.90	0.72	0.83
95% Confidence Interval	1.52%	1.29%	1.84%
Maximum Drawdown	2.0%	2.0%	2.0%
Longest Underperformance	38	78	44
Sterling Ratio	0.78	0.54	0.86
Sortino Ratio	1.24	1.04	1.18

Notes: Performance data assume an underlying 2% annual maximum drawdown target, generated by a naive buy-sell trading rule that is long (short) currency i versus the dollar whenever the forecast 4-week spot exchange rate is above (below) the 4-week forward rate. Weights for each portfolio indicate the implied weekly position in each exchange rate and, by aggregation, weekly total portfolio positions relative to a neutral (strategic) currency benchmark. Portfolio I weights the three exchange rate models together on the basis of return and risk correlations (that is, within standard Markowitz mean-variance optimisation framework). Portfolio II combines the models on the basis of risk minimisation. Portfolio III weights combines models on the basis of risk and drawdown characteristics of simulated returns, and using the constraint in equation (56) above, with β set at 0.5. All portfolios constructed using a Visual Basic optimisation tool. Transaction costs assumed to be 5 basis points round trip, based upon an informal survey of market participants.

Portfolio I Weights: 3.03 Euro; 0.95 Yen; 0.88 Sterling.
Portfolio II Weights: 1.00 Euro; 1.68 Yen; 2.45 Sterling.
Portfolio III Weights: 2.44 Euro; 1.83 Yen; 2.84 Sterling.

Table 20: Performance Data for Augmented Trading Strategy - Efficient Portfolios with Annual Return Target

	Portfolio		
	I	II	III
Annualised Return	5.00%	5.00%	5.00%
Annualised Risk	4.26%	8.97%	6.38%
Information Ratio	1.17	0.56	0.78
95% Confidence Interval	3.71%	7.81%	5.56%
Maximum Drawdown	10.40%	24.95%	16.57%
Longest Underperformance	165	169	168
Sterling Ratio	0.48	0.20	0.30
Sortino Ratio	1.73	0.85	1.18

Notes: Performance data assume an underlying 5% annualised return target, generated by a naive buy-sell trading rule that is long (short) currency i versus the dollar whenever the forecast 4-week spot exchange rate is above (below) the 4-week forward rate. Weights for each portfolio indicate the implied weekly position in each exchange rate and, by aggregation, weekly total portfolio positions relative to a neutral (strategic) currency benchmark. Portfolio I weights the three exchange rate models together on the basis of return and risk correlations (that is, within standard Markowitz mean-variance optimisation framework). Portfolio II combines the models on the basis of risk minimisation. Portfolio III weights combines models on the basis of risk and drawdown characteristics of simulated returns, and using the constraint in equation (56) above, with β set at 0.5. All portfolios constructed using a Visual Basic optimisation tool. Transaction costs assumed to be 5 basis points round trip, based upon an informal survey of market participants.

Portfolio I Weights: 3.44 Euro-Dollar; 1.68 Yen-Dollar; 0.00 Sterling-Dollar.
Portfolio II Weights: 2.62 Euro-Dollar; 4.20 Yen-Dollar; 7.84 Sterling-Dollar.
Portfolio III Weights: 3.41 Euro-Dollar; 2.42 Yen-Dollar; 4.28 Sterling-Dollar.

Table 21: Performance Data for Augmented Trading Strategy - Efficient Portfolios with Annual Drawdown Target

	Portfolio		
	I	II	III
Annualised Return	0.96%	0.40%	0.60%
Annualised Risk	0.82%	0.72%	0.77%
Information Ratio	1.17	0.56	0.78
95% Confidence Interval	0.71%	0.63%	0.67%
Maximum Drawdown	2.0%	2.0%	2.0%
Longest Underperformance	165	169	168
Sterling Ratio	0.48	0.20	0.30
Sortino Ratio	1.73	0.85	1.18

Notes: Performance data assume an underlying 5% annualised return target, generated by a naive buy-sell trading rule that is long (short) currency i versus the dollar whenever the forecast 4-week spot exchange rate is above (below) the 4-week forward rate. Weights for each portfolio indicate the implied weekly position in each exchange rate and, by aggregation, weekly total portfolio positions relative to a neutral (strategic) currency benchmark. Portfolio I weights the three exchange rate models together on the basis of return and risk correlations (that is, within standard Markowitz mean-variance optimisation framework). Portfolio II combines the models on the basis of risk minimisation. Portfolio III weights combines models on the basis of risk and drawdown characteristics of simulated returns, and using the constraint in equation (56) above, with β set at 0.5. All portfolios constructed using a Visual Basic optimisation tool. Transaction costs assumed to be 5 basis points round trip, based upon an informal survey of market participants.

Portfolio I Weights: 0.66 Euro-Dollar; 0.32 Yen-Dollar; 0.00 Sterling-Dollar.
Portfolio II Weights: 0.21 Euro-Dollar; 0.34 Yen-Dollar; 0.63 Sterling-Dollar.
Portfolio III Weights: 0.41 Euro-Dollar; 0.29 Yen-Dollar; 0.52 Sterling-Dollar.

4 The Role of Order Flow in Exchange Rate Forecasting

4.1 Introduction

While empirical evidence over the post-Bretton Woods period suggests that fairly standard macroeconomic fundamentals–such as relative monetary velocity–may influence the long-run behavior of real and nominal exchange rates (for surveys see Frankel and Rose, 1995; Froot and Rogoff, 1995; Taylor, 1995; Sarno and Taylor, 2002), modeling–and especially forecasting–the exchange rate over shorter horizons remains an occupational hazard of the international financial economist, since standard economic fundamentals appear to be poorly correlated with higher frequency exchange rate movements. Largely motivated by this stylized fact, a growing literature on market microstructure has emerged in recent years to suggest that the quality of fundamental-based exchange rate forecasts can be improved by resort to measures of foreign exchange order flow (defined as signed transaction volume: Evans and Lyons, 2002a, b; Lyons, 2001; Froot and Ramadorai, 2001), as well as variables such as surveys of market sentiment or positioning (Merrill Lynch, 2003).[48]

Order flow is initiated for a variety of reasons that differ across the various participants in the foreign exchange market. These participants include corporations, central banks, asset management firms, commodity trading advisors (CTAs), hedge funds, private individuals and investment bank dealers. Participants exhibit significant heterogeneity, in terms of opportunity sets and risk-return expectations, and display distinct informational asymmetries, with some participants better informed than others. By reputation, customer order flow is the primary source of private information in the foreign exchange market (Lyons, 1995; Ito, Lyons and Melvin, 1998; Bjønnes and Rime, 2003; Rime, 2001). This private information is typically assumed to relate to future innovations in fundamental exchange rate determinants, including monetary policy innovations (Evans and Lyons, 2002a; Lyons, 2003; Jansen and de Haan, 2003). But it can also incorporate knowledge of the decision-making process that triggers strategic shifts in portfolio benchmark hedge ratios in response to changes in risk appetite or return objectives independently from innovations in published fundamentals (Lyons, 2002a). Similar to innovations in fundamental variables, changes in long-term investment objectives will lead to asset allocation shifts within investment portfolios, for instance between international bonds and eq-

[48] Foreign exchange order flow should not be confused with transaction volume; the latter is a measure of trading activity between customers and dealers, or within the interdealer market, over a given period and in a particular exchange rate without indication of the direction of these transactions. By contrast, order flow is defined as signed transaction volume (Lyons, 2001), with the sign of a transaction determined by the initiating agent. Order flow therefore provides an indication of the relative strength of buy (sell) orders between, say, customers and dealers, with a purchase (sale) by the customer recorded as a net buy (sell). In this way, order flow within particular investor groups will not necessarily sum to zero, but can instead exhibit persistent trends if, say, customers build a long (short) position in a particular exchange rate relative to an underlying neutral benchmark position.

uities, that in turn inspire order flow.

A central hypothesis of the microstructure literature is that order flow allows the wider market to learn about the private information and trading strategies of better informed participants, and therefore represents the conduit through which informational asymmetries become embedded within market prices (Lyons, 1993; Bjønnes and Rime, 2001a). This hypothesis seems an intuitive explanation of the process of price discovery in the foreign exchange market. If valid, it implies that customer order flow will consistently be more important to the determination of exchange rate returns than interdealer order flow. In addition, order flow generally should have greater explanatory and predictive power for exchange rate returns than fundamental variables.

This paper seeks to make two main contributions. First, as a foundation to our empirical analysis we provide an extensive description of the structure of the foreign exchange market. This description focuses upon both the interaction of the main market participants and the current market infrastructure, and in our opinion represents the most comprehensive and accurate description of the foreign exchange market available. Second, with this foundation in place, and using aggregated and disaggregated customer order flow data from two major investment banks as well as the data on interdealer order flow employed by Evans and Lyons (2002a),[49] we critically evaluate the practical value of order flow data in terms of the accuracy of derived out-of-sample exchange rate forecasts. To our knowledge this is the first study that has assessed the practical value of foreign exchange order flow data under realistic trading conditions and using data available to market participants on a real-time basis.

The remainder of the paper is organised as follows. In the next section we discuss the structure of the foreign exchange market, focusing upon the key participants in the market and the nature and extent of their interaction. We then present a review of the market microstructure literature. The subsequent section presents our empirical analysis, and the final section draws conclusions from this analysis and provides suggestions for future research.

4.2 The Foreign Exchange Market

The foreign exchange market is the most liquid financial exchange in the world. Daily market turnover is estimated at approximately $1.9 trillion, including spot, forward and swap transactions (BIS, 2004). Including only spot transactions, daily turnover is approximately $621 billion. This compares to daily turnover in the US government bond market of $402 billion (Federal Reserve Bank of New York, 2003), and on the New York Stock Exchange of $40 billion (NYSE, 2003). London remains the major trading centre for foreign exchange,

[49] This database is available at http://faculty.haas.berkeley.edu/lyons/evanslyons.xls

[49] We thank Jan Loeys and Mustafa Caglayan at JPMorgan Chase and Peter Eggleston at the Royal Bank of Scotland for their help in obtaining the customer order flow data analysed in this paper. These data were pre-filtered and collated into indices by JPMorgan and the Royal Bank of Scotland to ensure that individual customer trades are not indentifable from the data, thereby maintaining customer confidentiality.

with 32% of daily market turnover transacted in this location (BIS, 2004). There are other important centres in New York and Tokyo, and smaller ones in Auckland, Sydney, Singapore, Hong Kong, Frankfurt and San Francisco.

In generic terms, there are two types of participants in the foreign exchange market, dealers and customers[50]. Dealers contribute the majority of market liquidity, with approximately 53% of total daily market turnover occurring within the interdealer market (BIS, 2004). Foreign exchange customers contribute the remaining 47%. Of this figure, financial investors contribute the majority share, equivalent to 33% of total market volume, and corporations 14%. Although a detailed breakdown of financial flows is not publicly available, informal discussions with a number of major investment banks suggest that asset management firms and hedge funds each account for approximately one quarter of customer flows, or 12% of total market flows each; CTAs, central banks and individuals contribute the remaining 9%[51].

Despite high liquidity[52], the foreign exchange market appears to be inefficient– in the sense that there are significant deviations from covered interest rate parity (Taylor, 1995; Sarno and Taylor, 2002)– and opaque, in the sense that the lack of a physical market place or market places makes the process of price-information interaction particularly difficult to understand (Dominguez, 1999; Lyons, 2002b), especially in comparison with either equity or fixed income markets. Market inefficiency implies the presence of persistent profit opportunities from informed trading (Baldridge, Meath and Myers, 2000). Although the introduction of the euro in 1999 encouraged a reduction in market opacity by replacing first eleven and subsequently twelve currencies with a single currency, other recent infrastructural developments have moved in the opposite direction; these include the increasing dominance of electronic interdealer trading that assures participants of ex ante anonymity (Portes, 2002).

Market opacity implies that information takes time to be fully reflected in prices. It also means that the process by which new information is embedded in exchange rates is unclear. Some researchers contend that the missing piece of this puzzle could be order flow (Lyons, 2001), and particularly customer order flow (Bjønnes and Rime, 2001b). Indeed, by reputation customer order flow is the primary source of private information in the foreign exchange market. Consequently, it is important to understand exactly what order flow data are measuring.

[50] Dealers are also known as broker-dealers. The two terms are synonymous.

[51] Market share data can vary substantially between investment banks who focus upon different segments of the customer market. Also, categorisation of flows between the various customer groups is also somewhat arbitrary, with a number of investors spanning more than one segment. Consequently, these detailed market share data should be interpreted with caution.

[52] Consistent with Kyle (1985), we define a liquid market as one that exhibits the following characteristics: tightness, so that bid-ask spreads for small transactions are tight; depth, so that bid-ask spreads for large transactions are small and; resilience, so that deviations of the spot rate from Fair Value should be corrected quickly. See Daníelsson and Payne (2001) for a recent assessment of these criteria applied to the interdealer sector of the foreign exchange market.

Traditional asset-price models of exchange rate determination take the form:

$$S_t = \beta' F_t + \alpha(E_t S_{t+1} | I_t), \tag{56}$$

where S_t is the spot exchange rate at time t, and $(E_t S_{t+1} | I_t)$ is the spot rate expected at time $t+1$ given information at time t (both expressed as domestic currency per unit of foreign currency). F_t represents a vector of fundamental variables that form a cointegrating relationship with, and exhibit some persistent explanatory power for the determination of exchange rates, and β represents a vector of factor loadings. Following Frankel and Froot (1985) and Froot and Frankel (1989), researchers have often measured exchange rate expectations by using some proxy series such as survey data of market expectations. Available measures include Consensus Economics (Consensus Economics, 2004). However, the predictive ability of survey data is typically low, and often Granger-caused by spot exchange rates (Brown and Maital, 1981; Taylor, 1988). Thus, investor survey responses react to, rather than preempt exchange rates moves.

In principle, order flow data circumvent the necessity to proxy expectations by directly measuring the activity, and by implication the expectations, of foreign exchange market participants (Evans and Lyons, 2002a). But while customer order flow may represent the missing piece of the exchange rate puzzle–we aim to test this hypothesis in this paper–data on customer order flow are only contemporaneously observable on a real-time basis by a limited number of well-informed market participants. These participants include the custodian or investment bank or electronic trading platform that collects and collates the data, and–on an ad hoc verbal basis only–a select group of preferred customers of these organisations, for instance large asset management firms, CTAs and hedge funds[53]. This limited access to order flow data reflects confidentiality concerns of major market participants whose trading activity is captured by these data and which could therefore be identified by competitors, and a wish on the part of the collecting institution to sustain and profit from this potential information advantage for as long as possible. For this reason, more systematic access to proprietary order flow data by preferred customers is provided only after the raw flow data have been filtered and collated into an index to ensure customer anonymity (for instance, see the description below of the customer order data made available to this study), and with at least some publication lag. We discuss below the extent to which pre-filtering and publication lags reduce, from a customer perspective, the information content of available order flow data.

By contrast, dealers in small banks see most customer order flow data only indirectly once they have been embedded in prices. Furthermore, most customers do not gain direct access to even filtered flow data, and instead have

[53] Preferred clients are those identified by investment banks to be of strategic importance to their position in the foreign exchange market. As well as best execution, these clients will be provided with better streaming information than other clients on intraday and daily order flow and market positioning, and will be allocated resources for collaboration on confidential, customised research. It is important to emphasise that information concerning order flow data made available to preferred customers never includes details of individual customer transactions, so that customer anonymity is assured.

to content themselves with interpretative analysis provided by dealers that includes qualitative trading strategies based upon developments in their own collated flow data (HSBC, 2003; Citibank, 2003). We are not aware of any rigorous quantitative evidence to suggest that this interpretative approach can add value to investment portfolios on a persistent basis, once publication lags, representative trading costs and interest carry have been incorporated into back-test simulations.

In a later section of this paper, we empirically test the reputation of customer flows as the primary source of private information in the foreign exchange market. Before doing this, it is important that we describe the main participants in the foreign exchange market, and the nature and extent of their interaction. In a fast evolving market, we believe that this description is both the most comprehensive and current available.

4.2.1 Market Microstructure

Theoretical Model The academic literature presents a stylised model of the foreign exchange market, and of the interaction between the various customer groups and dealers. The most recent characterisation is due to Evans and Lyons (2002a,b), which in turn borrows aspects of Kyle's (1985) sequential equilibrium model. It describes the foreign exchange market as a decentralised dealership (Lyons, 2002a). There are N dealers in the market, a continuum of non-dealers (or customers), and an infinite number of trading days. Dealers observe a periodic payoff on foreign exchange, denoted R_t, representing the flow of macroeconomic information. This payoff is proxied by Evans and Lyons by interest differentials. There are three stages to the foreign exchange trading day. First, and having observed R_t at the start of each day, dealers independently and simultaneously set bid-ask spreads for, and trade with, customers. Second, dealers trade amongst themselves, independently and simultaneously posting a bid-ask spread for other traders. These spreads lead to trades, as dealers spread risk generated by earlier customer trades through the interbank market. Once this second round of trading is complete, all dealers are able to observe the order flow, X_t, that has occurred from interdealer trading during that day. These data are assumed by Evans and Lyons (2002a,b) to convey important information about the size and sign of customer trading during Round One. Finally, in the third round of trading dealers once more trade with customers, in order to share overnight risk more widely across the market. A crucial assumption here is that dealers set prices in Round Three such that customers willingly absorb all dealer inventory imbalances, so that dealers run no overnight open positions. Consequently, the closing price at the end of each day within the Evans and Lyons model will be

$$P_t = \beta_1 \sum_{\tau=1}^{t} \Delta R_t + \beta_2 \sum_{\tau=1}^{t} X_t \tag{57}$$

and the change in prices from the end of day $t-1$ to the end of day t can be written as

$$\Delta P_t = \beta_1 \Delta R_t + \beta_2 X_t \tag{58}$$

where $X_t = \sum_{i=1}^{N} T_{it}$, is the total order flow generated by interdealer trading during the day and $T_{it} = \alpha C_{it}^1$ is the sum of customer orders received by individual trader i during the first round of trading. α is a constant coefficient.

The Evans and Lyons (2002a,b) characterisation of the foreign exchange market is one that few practitioners will recognise. In particular, the notion that customers willingly absorb the daily inventory imbalance of dealers is implausible. Indeed, customers have traditionally paid dealers to assume this type of price and credit risk on their behalf, allowing them to access the liquidity of the interbank market through only a few contact points rather than having to establish individual credit agreements with every member of the dealer community[54]. This payment is made in soft terms, with dealers traditionally been able to divide customer trades into smaller tranches as they spread risk through the interbank market. As the width of bid-ask spreads is a positive function of transaction size (Bjønnes and Rime, 2001b), tranching saves the dealer transaction costs and, along with knowledge of, and trading around customer limit order books, has traditionally represented a major source of trading profit for banks (Rime, 2001).

This profit source has recently been undermined by the behaviour of informed customers increasingly aware of the strategic price impact of their trading activity. Much of the skill of a customer-based trader derives from his ability to minimise the extent of market chatter surrounding his trading activity. To this end, for example, many currency managers will now regularly engage in tranching themselves, dividing trades of \$100 mn-\$1 bn,[55] into smaller amounts of, say, \$25-\$50 mn. These tranches are then spread amongst interbank dealers intermittently during the course of a trading session via a number of dealers. Although consistent with customer profit-maximisation, this practice undermines the trading profits of dealers and implies that banks are now accepting risk on behalf of customers without adequate compensation in return. How dealers can gain payment for this risk in the future is not immediately apparent, given the rise of electronic trading that implies both a compression of dealer bid-ask spreads and an increase in ex ante customer anonymity (Portes, 2002). Consequently, risk sharing in the foreign exchange market would seem to be an important area of future research.

Practical Aspects As the existing theoretical model presents a stylised and simplistic view of the foreign exchange market, it is important to describe in

[54] For instance, it is standard practice for a currency overlay manager (the agent) to request prior approval from a client (the principal) to trade with at least ten dealers, on a continuous basis on the client's behalf. This ensures that the price and credit risk of the client with respect to any one of the dealers is minimised and that the overlay manager maximises its opportunity to achieve best execution for the client.

[55] \$1 bn is referred to as a "yard" in foreign exchange slang.

detail the types of participants that coexist in the market, the manner of their interaction and the associated market infrastructure.

Interdealer Market

The interdealer market encompasses market-makers, leverage traders, designated proprietary ("prop") traders, and senior risk takers. Market-makers continue to perform their traditional core function of facilitating access for customers to interdealer liquidity and providing best execution for customer trades. But their wider role has changed over recent years, in a number of ways. First, market-makers are typically allocated a book exchange rate upon which they focus their attention. This compares with a few years ago when these individuals traded in a range of exchange rates. Second, Danfelsson and Payne (2001) provide evidence to suggest that market makers often now focus upon one side of the market at a time, rather than posting genuine two-sided quotes. Put another way, market-makers are no longer typically the main source of price discovery in the foreign exchange market and are not attempting to generate excess profits from their market activity, but are instead largely facilitators of customer trades.

The traditional academic characterisation of designated prop traders is as intra-day, or even "nintendo", traders (Bjønnes and Rime, 2003), whose investment time horizon extends from minutes to hours at most, and for whom there is no capacity to run overnight positions. In fact, this description is more consistent with the activities of leverage traders[56]. These individuals trade primarily on the basis of order flow executed by the bank's trading desk, and typically have an investment horizon of a few hours, or at most days. Accordingly, the short-term nature of these traders' activity represents a key source of total market volatility, with tight stop-loss levels typically introduced around every position. As a result, price breaks of key technical levels typically extend further in the short term than they otherwise would in the absence of leverage interdealer trading (Osler, 2002, 2003).

Prop traders are actually often positively discouraged by senior management from trading too actively and instead are encouraged to focus most of their allocated risk budget upon longer investment horizons than leverage traders, typically days but sometimes up to three months or even one year. Risk budgets are reduced overnight, reflecting the difficulty of monitoring positions outside of business hours, but remain positive.

In addition, and previously overlooked by academic analysis and surveys of the foreign exchange market, are senior risk takers at large investment banks. These individuals perform a similar function to designated prop traders, but are allocated a much larger risk budget, for instance \$100-\$200 million, compared with \$25-\$40 million for designated prop traders, reflecting their relative seniority, experience and performance track record within the market[57]. This budget

[56] These individuals are also known as spot traders.

[57] Our estimate of the typical prop trader risk budget is smaller in absolute terms than Cheung and Chinn (1999), who studied the US sector of the market, and Cheung and Wong (2000), who examined trading activity and practices in Hong Kong and Singapore. But it is directionally consistent with both of these studies given broad market trends described below.

will also be reduced overnight. Senior risk takers are similarly expected to focus upon relatively long investment horizons.

A number of infrastructural trends are apparent within the interdealer market. First, industry consolidation has meant that the number of banks that account for a majority of interdealer flows has fallen since 1998, with 17 banks in London and 13 banks in the US accounting for 75% of turnover transacted in these locations. This compares with 24 and 20 banks in 1998, respectively (BIS, 2002). Second, the general level of risk appetite within the interdealer market has declined in the wake of the 1998 Long-Term Capital Management crisis, leading to a reduction in the level of risk capital allocated to individual traders. Third, and related to the previous point, the rigour of risk management procedures and infrastructure related to dealer activity has improved substantially in recent years, to a high standard. The predominant focus of risk control procedures remains the imposition of maximum intraday and overnight nominal (dollar) position limits for individual traders. Many banks also make the amount of risk capital available to traders a function of past performance, rewarding good performance with an increase in risk capital and penalising bad results. Typically, traders assume responsibility for, or "wear" in market jargon, all losses, so that these reduce next month's available risk capital on a commensurate basis. By contrast, profits are typically shared with the bank, such that risk capital available to a trader in the next month will increase only by some fraction of last month's reported profits. Trading risk is also monitored using a variety of other metrics, including daily maximum drawdown, or capital loss, and daily Value at Risk (VaR) limits[58]; the use of VaR limits appears now to be more commonplace than reported by Cheung and Wong (2000).

A loss that trips intraday limits is unlikely to signal the termination of a trader's risk budget, but instead will trigger notification of senior management, require a detailed explanation of the circumstances surrounding the loss and of remedial steps to be taken, and closer scrutiny of the trader's outstanding positions and activity in the immediate future. More generally, banks also conduct VaR sensitivity analysis on the activity of its trading desk in aggregate to ensure that senior management have a broad understanding of the probability of an extreme event that could threaten the solvency of the institution.

Customers

Customers interact with dealers in order to access the liquidity of the interbank market. By reputation, customer order flow is the most important source of private information in the foreign exchange market (Lyons, 1995, Ito, Lyons and Melvin, 1998, Bjønnes and Rime, 2003), with the sign of customer trades considered more informative than the associated nominal value of these trades (Bjønnes and Rime, 2001b). An important reason for this reputation is the het-

Exact position limits are considered to be market-sensitive information, and are therefore confidential to individual banks. Our estimates result from informal conversations with market participants.

[58] Value at Risk (VaR) is defined as the maximum percentage value of an investment portfolio that could be lost during a fixed period (e.g. one day) within a certain confidence level (e.g. 95%).

erogeneity that exists within this segment of the foreign exchange market. The term customer embraces corporations, central banks, asset management firms, CTAs, hedge funds and individuals. Asset management firms in turn incorporate fixed income and equity investors, but also currency overlay managers.[59] [60] There are a number of aspects to customer heterogeneity.

Opportunity sets and risk/return preferences

The various customer groups exhibit very different opportunity sets and risk-return preferences[61]. The market incorporates a relatively small set of active, and informed, customers - particularly currency overlay managers, hedge funds and CTAs - that exploit persistent profit opportunities by implementing tactical currency positions into portfolios around strategic benchmark exposures (Baldridge, Meath and Myers, 2000, Knott, 2002).

Many customers are passive, and uninformed.[62]. These customers initiate order flow primarily to pay and hedge foreign currency cost and revenue streams generated by international purchases and sales. It is typically neither price dependent nor initiated explicitly to turn a profit, but instead is intended to minimise the translation risk associated with foreign exchange exposure and volatility. In a similar vein, many asset management firms do not seek to maximise profits from foreign exchange trading, but instead passively hedge exchange rate exposure inherited from the sale and purchase of underlying assets back to a strategic currency hedge ratio without tactical consideration of the exchange rate at which these transactions are conducted (Dalio, 2002).

Although an increasing number of central banks allocate a small part of official foreign exchange reserves to wealth creation (Thomson, 2002), reflecting both an increase in the stock of reserves, particularly in Asia, and less need for these reserves to be held in liquid form due to wider use of derivative instruments, profit-maximisation remains a secondary issue. Indeed, as Dalio (2002) reports, in aggregate the major central banks have lost money on foreign exchange intervention during the floating exchange rate era. Nonetheless, reflecting the magnitude of associated flows central bank foreign exchange activity represents arguably the most keenly sought source of private informa-

[59] In a currency overlay programme, management of the underlying international asset exposure within a portfolio is separated from the management of associated currency exposures (European Central Bank, 2003). Tactical management of currency exposure around a strategic benchmark by specialist currency managers can provide investors, for instance corporate pension plans, with an additional source of diversified portfolio return (Baldridge, Meath and Myers (2000).

[60] In foreign exchange jargon, asset management firms are known as "real money". The term is used to indicate that financial or real assets, including equities, bonds or real estate, underlie the foreign exchange activity of these customers. CTA and hedge fund flows are known as "speculative money", as associated currency orders are based upon notional capital portfolios. Increasingly, this distinction is becoming blurred with a number of real money managers now offering leveraged currency funds based upon notional capital portfolios.

[61] There are similarities in the categorisation of customers in this section with Kyle (1985). He defines a market in which three types of participants coexist: nondiscretionary liquidity, discretionary liquidity and informed traders.

[62] Whether a customer is passive or active is determined not by the group from which they are drawn - for example, corporations versus asset managers - but by the nature of their activity in the foreign exchange market.

tion.[63] This information can relate either to prospective market intervention or compositional changes in the currency denomination of official foreign exchange reserves. For instance, on a number of occasions since the inception of the euro in January 1999, and particularly during 2003, market anecdote suggested important portfolio adjustments away from the US dollar and towards the euro amongst Middle Eastern and Asian central banks. As such activity occurs, transacting dealers gain important information that can be used to inform their own foreign exchange trading strategies, at least in the short term.

Investment Styles

Heterogeneity can also be observed in the range of investment styles that coexist within active - that is, informed profit maximising - customer groups. These styles exploit different types of information, leading to differences in the way that customers interpret public information announcements, and therefore differences in exchange rate expectations.

Many CTAs are pure technical, or "black box", managers. The associated investment process will typically comprise a set of optimised technical, or chartist, trading rules that have no intuitive underlying theoretical economic interpretation[64]. News to this customer group is historical price innovation, over any period from minutes, hours, days, all the way out to years, combined with detailed trading rules related to key support and resistance levels, moving average cross-over levels, over-bought and over-sold calculations[65] and a range of other price patterns. Publicly announced macroeconomic news is only relevant indirectly, to the extent that it has some historic price impact. The importance of this type of technical trading within the foreign exchange market, particularly for high frequency exchange rate returns, is confirmed by Taylor and Allen (1992), Cheung and Wong (2000), and Euromoney (2002).

By contrast, hedge funds and currency overlay managers initiate order flow predominantly on the basis of publicly available macroeconomic information, with an expected pay-off schedule that is likely to stretch from around one to three months. The investment process of these customers is often highly quantified, with pre-defined trading rules based upon theoretical relationships between economic or financial variables and exchange rates. In this case, public data innovations will directly trigger customer order flow.

Some hedge funds and asset managers also introduce tactical exchange rate hedges into portfolios on the basis of a purely qualitative interpretation of data and technical price patterns, on the rationale that not all events-including central bank intervention-that determine exchange rate returns can be quantified in a consistent manner over time.

[63] Total central bank foreign exchange reserves are estimated at $2,500 billion, up from $1,000 billion in 1990 (Thomson, 2002).

[64] Some CTAs will adopt a more discretionary investment style. Models will be used to determine key technical levels and turning points but positions will be implemented on a discretionary basis by portfolio managers and traders.

[65] Over-bought and over-sold calculations attempt to define when an exchange rate has moved too far and fast in either direction. They are typically calculated based on a moving average of the difference between the number of advancing and declining days over a certain period of time.

Finally, risk control currency managers introduce option replication strategies into client portfolios in order to minimise the downside risk attached to any level of foreign exposure over some fixed investment horizon, typically one year (Layard-Liesching, 2002). Consequently, this type of manager will react to price and data innovation indirectly to the extent that these affect the downside risk profile implicit within portfolios.

Reaction Speeds

Heterogeneity also implies differences in the reaction speed of customers to the arrival of new information, with some managers more nimble than others. In a contradiction of the Rational Expectations Hypothesis (REH) that argues public information is instantaneously reflected in exchange rates, this form of heterogeneity is consistent with the practical observation that the contemporaneous relationship between order flow and exchange rates is relatively persistent (Froot and Ramadorai, 2001; Fan and Lyons, 2001; Evans and Lyons, 2002a, b; Bjonnes and Rime, 2001a).

Differences in reaction speeds can provide dealers with important information about the size and sign of customer trades, and the likely persistence of this trading. To the extent that a dealer sees order flow from large informed customers, it is generally reasonable to assume that this is only a small part of the total trade being executed, and that the remaining orders are likely to be fed into the market throughout the trading session. This knowledge will allow the dealer either to "piggy-back" on the trade, committing some of his own risk capital to the same trade (Bjønnes and Rime, 2001b), or to net off trades from other customers.

Information about large customer order flow is made available by dealers to other large customers, on a quasi-anonymous basis via direct voice links, once a dealer has executed a customer order. An indication of the exchange rate, type of initiating customer, and size and sign of the transaction will be provided, as well as the ease with which the market absorbed the order.[66] This information helps other customers gauge the extent of technical support and resistance levels around the current spot exchange rate, including any option-based trading structures, and the probability and likely extent of any break-outs from prevailing price trend channels.

In a similar manner, ad hoc information concerning the level and volume of guaranteed stops placed by customers on investment order books will be provided by dealers to preferred clients on a daily basis. Osler (2002) finds that information regarding stop orders has historically had an important explanatory role for yen-dollar. But the quality of this information is likely to have deteriorated in recent years as customers have become more reticent to place stop orders with dealers, reflecting greater cogniscence of their own strategic market impact. For informed customers, this cogniscence has increasingly focused efforts upon ensuring that internal risk management systems can incorporate appropriate, continuous position gain/loss monitoring procedures to allow creation and execution of trades as stop levels are approached.

[66] The dealer will never provide the name of customers initating trades to other customers.

Market Infrastructure Most interbank trading occurs electronically, via either the Electronic Banking System (EBS) or Reuters D3000. Both systems were established in 1993 and were the primary facilitators of the subsequent marked increase in market liquidity. Their functionality is essentially equivalent, providing ex ante anonymous limit order bid-ask pricing to dealers. Combined, these systems account for approximately 85% of total interbank activity, with EBS dominating in all exchange rates other than sterling, Canadian and Australian dollar cross rates[67]. The remaining interbank trading activity is shared by the remnants of voice broker (10%) and secure bank-to-bank chat lines provided by Reuters (5%).

Voice trading continues to account for the majority of customer trades. But three electronic systems - FX Connect, FXAll and Currenex - have recently been introduced to the customer-dealer space[68]. All three systems are multi-bank electronic trading portals linked directly to customer trading desks. They comprise an essential component of the push towards the introduction of Straight Through Processing (STP) that facilitates the complete automation of customer foreign exchange management from order creation through electronic trading portals to the settlement and confirmation of trades. A primary motivation behind the introduction of STP is a reduction in the risk of human error at various points in this process. For a survey of recent foreign exchange e-trading and STP developments at a major currency overlay manager, see Baird (2002).

FX Connect and FXAll transact approximately equivalent daily volumes[69]. But the greater automaticity of FXAll ensures that transactions are more efficient through this system, minimising the risk of trading errors and better facilitating the aggregation of client trades and netting of risk by dealers. Accordingly, these facets imply that this portal will grow in relative importance over time. FXAll provides customers with automatic streaming bid-ask quotes on request - so-called Requests for Quotes, or RFQ's - simultaneously from a number of dealers in any size of transaction, and for most traded exchange rates[70]. Details of any subsequent deals transacted in FXAll are confidential to the customer and transacting bank alone, and are not seen by other banks given an RFQ. This contradicts Evans and Lyons (2002a), who suggest that the shift away from voice trading and towards electronic portals will increase the availability of order flow data, and thereby improve the transparency of the foreign exchange market and the quality of microstructure-based exchange rate models. In reality, electronic order flow is treated as strictly confidential by

[67] Why this delineation by exchange rate originally arose is not clear. Unsuccessful attempts have periodically been made by both systems to gain market share in additional exchange rates.

[68] A fourth system, Atriax, closed in 2002.

[69] At the time of writing, FX Connect and FXAll transact an estimated $11 bn of customer orders per day, and Currenex between $4.5 and $6.5bn (E-forex, 2003, FXAll, 2003).

[70] Market anecdote suggests that the optimum number of quoting banks is between three and five. This choice ensures that quotes are competitive, but also minimises dealer knowledge of customer activity. Quotes on all three systems are market quotes, with no limit order facility available on either system. Consequently, customer limit orders are conducted by voice links.

system governing boards[71].

FX Connect is primarily an 'At Best' trading platform, meaning that customers approach individual dealers via linked computer terminals to provide a two-way quote for a specified transaction size at the best price possible. Once a quote is received, the customer can choose to do one of three things: accept the quote and trade at that price with the quoting dealer, reject the quote and approach a second dealer, or to terminate the transaction altogether. Either way, the information discussed by customers and dealers is treated as strictly confidential and, again, is not published more widely.

The structure of Currenex is similar to FXAll. But unlike the other two systems that require dealers to pay a fee in order to quote prices, Currenex levies a charge on customers for each trade transacted on its portal. This differentiates the type of customer using the various systems, with Currenex customers largely confined to hedge fund and corporate clients.

4.3 Literature Review

A number of empirical studies have demonstrated an explanatory role for order flow in exchange rate models. Most of these studies have focused–principally for reasons of data availability–upon the role of the interbank market, implicitly maintaining the assumption that these flows embed information into prices by reacting to customer orders, and that customer orders are the primary source of private information in the foreign exchange market. As Taylor (2003) notes, this assumption implies a substantial–and perhaps implausible–multiplier effect from (informed) customer to interdealer orders, given the relative importance of these two groups in total daily turnover. We assess the validity of this assumption in this study.

In a path-breaking study of foreign exchange market microstructure, Evans and Lyons (2002a) analyse the ability of interdealer order flow data collected from Reuters D2000-1 to explain the daily variation of mark-dollar and yen-dollar during a four-month period from May to August 1996. To this end, Evans and Lyons regress the daily log change in each exchange rate on the change in interest differentials–their proxy for fundamentals–and daily interdealer order flow. They find that 64% of daily mark-dollar returns and 45% of yen-dollar returns can be explained within this simple framework. Moreover, on the basis of Wald exclusion tests, the explanatory power of these regressions is almost wholly due to order flow (Lyons, 2001).

In a subsequent paper, Evans and Lyons (2002b) extend this analysis to an additional seven exchange rates: the price of the US dollar expressed in terms of UK sterling, the Belgian, French and Swiss francs, Swedish krona, Italian lira and the Dutch guilder.[72] Contemporaneous correlations are more variable over this extended group, with R^2 statistics ranging from 0.00 for the Belgian franc

[71] However, the act of initiating RFQs with a number of banks on electronic portals may accentuate the price impact of transactions, simply because a number of brokers will be alerted to the presence of a potential flow.

[72] In this paper, Evans and Lyons exclude interest differentials.

and Swedish krona to 0.68 for the mark (slightly higher than in the dollar-mark regression of Evans and Lyons, 2002a)[73]. In addition, Evans and Lyons also report a substantial rise in explanatory power with the inclusion in equations of order-flow into other currencies. For instance, the estimated R^2 for the Swedish krona-dollar exchange rate increases to 0.69, and to 0.78 for mark-dollar. Although this type of cross-pollination effect of order flow between exchange rates seems intuitive–the foreign exchange market does experience general dollar buying and selling trends–the sign of estimated coefficients is unstable, switching from equation to equation. Thus, net buying by dealers of sterling enters with a positive coefficient into the equation for the Belgian, Swedish and French exchange rates against the dollar, but with a negative coefficient for the mark, yen, Swiss franc, lira and guilder.

The finding of a substantial contemporaneous correlation between interdealer order flow and exchange rate returns is confirmed by Fisher and Hillman (2003), albeit with generally lower R^2 statistics than Evans and Lyons (2002a, b), and Daníelsson, Payne and Luo (DPL, 2002). DPL examine Reuters D2000-2 data over a ten month period from 1999 to 2000 using eight sampling frequencies ranging from five minutes to one week for the mark-dollar, yen-dollar, sterling-dollar and sterling-mark. R^2 statistics for the mark-dollar and yen-dollar equations typically lie in the region of 0.3 - 0.5 across all sampling frequencies, except for very high frequency observations for yen-dollar which are closer to zero. Arguably, the analysis of sterling cross rates provides the most robust results, given the relative market share of Reuters D2000-2 in these exchange rates. Although DPL do find a significant role for order flow in explaining sterling returns, R^2 statistics are generally lower and diminish dramatically for frequencies beyond one hour, suggesting a much weaker relationship than for the other exchange rates.

DPL also consider the practical relevance of their results by assessing the out-of-sample predictive power of interdealer order flow for exchange rate returns. To this end, they provide a set of results termed "genuine forecasts", by which they mean forecasts that include no perfect foresight[74]. Largely consistent with our findings reported below, they conclude that without perfect foresight order flow has no predictive power for exchange rate returns: R^2 statistics fall to zero, root mean squared forecast errors (RMSFEs) are higher for forecasts based upon order flow than a naive random walk and, indeed, order flow is Granger-caused by exchange rate returns. Payne and Vitale (2002), on the basis of an examination of central bank order flow for Swiss franc-dollar over the sample period 1986 to 1995, similarly conclude that "the leads and lags of Swiss National Bank customer order flow [in Swiss franc-dollar] often have the wrong sign and are, in general, not significantly different from zero" (Payne and Vitale, 2002). These findings would seem to question the validity of an important central hypothesis of the microstructure literature.

[73] We assume that the R^2 statistic reported for Swiss franc-dollar (-0.53) is a typo.

[74] This differentiates their approach from the traditional Meese-Rogoff (1983a, b) methodology, and is therefore more representative of reality. DPL also report Meese-Rogoff forecasts. Unsurprisingly, these are more favourable to the microstructure literature.

Other researchers have reported the existence of a significant explanatory role for order flow in exchange rate movements. Using daily EBS data from January 1998 to December 1999, Killeen, Lyons and Moore (KLM, 2002) determine the presence of a Granger-causal relationship that runs from French franc-mark order flow to the exchange rate. They also test for the presence of a cointegrating relationship between the exchange rate, order flow and interest rate differentials, which, in common with Evans and Lyons (2002a), they include as a proxy for macroeconomic fundamentals. One can, of course, reasonably question the use of cointegration analysis on such a short data span, as the hypothesis embedded in this analysis is distinctly long term. This criticism notwithstanding, KLM find in favour of the existence of one cointegrating vector that allows them to infer the presence of long-term co-movement between these three series. Furthermore, the results change little with the inclusion of interest differentials, suggesting that the coefficient on this variable is zero. Mende and Menkoff (2003), using customer order flow for euro-dollar from a medium-sized (anonymous) German bank during the sample period July to November 2001, report similar findings from cointegration tests[75].

By contrast, Bjønnes and Rime (2001a), using weekly US Treasury data for interdealer order flow traded by major participants in the US sector during the sample period July 1995 to September 1999, report mixed results from cointegration tests[76]. They find in favour of a cointegrating relationship for mark-dollar, sterling-dollar and Swiss franc-dollar, but against for yen-dollar and Canadian dollar-US dollar. It is unclear why this difference exists, given that all of the exchange rates share the same numéraire, unless it is indicative of the problems implied by use of an inherently long-term hypothesis with a short data span. In addition, Bjonnes and Rime (2001a) conclude that inclusion of foreign equity returns generates a marked increase in R^2 statistics of estimated regressions, suggesting that it is primarily this variable, and not order flow, that has explanatory power for exchange rates. This is true also when the authors remove perfect foresight from regressions, by including lagged order flow as an explanatory variable for exchange rate returns; in this case, the explanatory power of order flow becomes insignificant for four of five exchange rates.

Overall, therefore, results from analysis of the relationship between aggregate interdealer order flow and exchange rate returns provide a number of conflicting results and raise a series of question marks over the central hypotheses of the market microstructure literature. In particular, none of the studies discussed above is able to demonstrate that order flow has predictive power for future exchange rate returns rather than simply contemporaneous correlation.

All of the preceding studies concentrate upon an analysis of interdealer order flow. They implicitly maintain the assumption that customer orders are the main source of dispersed private information in the foreign exchange market and that the interplay between customer and interdealer orders represents the

[75] Mende and Menkoff (2003) use tick data over 87 trading days.

[76] Bjønnes and Rime define a major participant as one with more than $50bn equivalent in foreign exchange contracts, including spot, forward and options, on the last business day of any quarter during the previous year (Bjønnes and Rime, 2001b).

mechanism by which this information is embedded in prices. A lack of available data has generally prevented direct analysis of customer order flow. An exception is Fan and Lyons (2001), who present a qualitative analysis of customer order flow disaggregated by customer type and transacted with Citibank; this institution accounts for approximately 10% of daily customer order flow (Euromoney, 2003). Fan and Lyons conclude that the price impact of order flow for yen-dollar and euro-dollar is differentiated by customer type, with real money flows more adept than speculative money flows at anticipating turning points in exchange rate trends. This finding contradicts the popular view of speculative investors as market leaders within foreign exchange (Cai, Cheung, Lee and Melvin, 2001). Evans and Lyons (2003), on the basis of a more quantitative treatment of Citibank order flow than Fan and Lyons, also conclude that the explanatory power of order flow differs significantly between the various customer groups. But apparent instability in estimated coefficients, with signs switching between positive and negative depending upon the exact equation specification, cautions against over-emphasis of these results.

Another variant of the microstructure literature considers the explanatory power of informed versus uninformed order flow for exchange rate returns. Ito, Lyons and Melvin (ILM, 1996) examine the role of informed order flow in the Tokyo foreign exchange market. They compare trading activity pre- and post-December 1994, the date of the introduction in Tokyo of previously prohibited lunchtime interdealer trading between noon and 1.30pm. Interdealer trading should encompass relatively informed trading as it incorporates knowledge of customer activity. ILM do indeed find a significant impact on the volatility of yen-dollar returns with this change to market structure, both in terms of an increase in volatility around lunchtime as well as in the general pattern of exchange rate volatility throughout the whole Tokyo trading session. Since the flow of public information was not affected by the change in market structure, ILM (1996) conclude that volatility shifts are synonymous with the presence of private information. By implication, informed Tokyo order flow has a predictive role for high frequency returns to yen-dollar.

By contrast, analysing the same data set, Anderson, Bollerslev and Das (ABD, 1998) find no evidence of a significant shift in volatility patterns caused by the change in market structure, although they do not necessarily exclude the possibility of an important role for private information in the foreign exchange market. ABD also question the efficacy of the variance ratio methodology of ILM (1996), suggesting that it will provide invalid results when applied to high frequency data reflecting the incorrect assumption of normally distributed exchange rate returns. More supportive of the findings of ILM (1996) are the conclusions of Daníelsson and Payne (2001), who examine Reuters D2000-2 order flow, bid-ask spread data and measures of order book depth for mark-dollar during one week in October 1997 and conclude that volatility shifts are associated with the presence of informed market participants in that exchange rate.

In a similar vein to ILM (1996), and using a similar data set, Covrig and Melvin (2001) establish the primacy of informed Tokyo-based Japanese traders in yen-dollar. As a result, the occurrence of a cluster of informed Tokyo-based traders generates yen-dollar quotes from Tokyo that lead quotes in this exchange rate from other trading centres, such as London or New York. Bjønnes, Rime and Solheim (2003) examine Riksbank volume trading data for the Swedish krona-euro[77] exchange rate between January 1995 and December 2001 and also conclude that the relationship between order flow and exchange rates depends upon the instigating institution. In particular, order flow instigated by large brokers that have maintained a local presence in the Swedish krona-euro market for an extended period has more price impact than the equivalent flow from relatively smaller banks that are less well established in the domestic market. Again, therefore, informed order flow appears to have a significant explanatory role for exchange rate returns. Andersen, Bollerslev, Diebold and Vega (2002) reach a similar conclusion using high frequency return data for a range of exchange rates over a six-year sample period. And Payne (2000), analysing Reuters D2000-2 data for mark-dollar, concludes that 60% of variation in bid-ask spreads can be explained by asymmetric, or private, information. The magnitude and direction of the impact on spreads from private information documented by Payne is consistent with evidence from Lyons (1993). Payne thus concludes that bid-ask spreads contain a significant adverse selection component, with dealers adjusting quoted spreads to reflect the probability that they are trading with an informed market participant.

In contrast to theoretical prediction, a number of studies also conclude in favour of a significant explanatory role for uninformed, or expected, order flow. For instance, while Bjønnes, Rime and Solheim (2003) conclude that approximately one third of the volatility of daily Swedish krona-euro returns can be explained by informed order flow, they also find that uninformed flow can explain an additional one fifth of daily returns.

These findings are consistent with the conclusions of Osler (2002, 2003), who analyses price quote data from Reuters over the sample period January 1996 to April 1998 with a bootstrap methodology to assess the significance of price cascades triggered by guaranteed stop-loss orders.[78] She finds in favour of a significant price impact from guaranteed stop-loss orders, and therefore concludes that non-informative order flow does exert a significant impact upon exchange rates. These findings are supported by Bates, Dempster and Romahi (2003), who examine the daily order book of another major investment bank-HSBC-during the sample period March to August 2002 using pattern recognition techniques derived from computational learning.

Another variant of the market microstructure literature assesses the issue of whether the strength of the relationship between order flow and exchange rates is dependent upon prevailing market conditions (Payne, 2000, Luo, 2001,

[77]Mark prior to January 1999.

[78]These are often also termed Open or limit orders. The terms are synonymous. These orders provide the instigating party with a guaranteed price at which the associated order will be filled by the dealer, are either stop-loss or take-profit, and will be buy or sell orders depending upon the type of underlying trade (buy versus sell). Anecdotal evidence suggests that the typical size of a customer stop order is $10 mn, and has an average duration of 2 days. Dealer stop orders have an average value of $2 mn and typically exist for just 2-3 minutes (Melvin & Wen, 2003).

Lyons, 1996, and Osler, 2002b). On the basis of Reuters D2000-2 data over a maximum sample period of September 1999 to July 2000, and for a variety of modelling techniques and exchange rates, Luo (2001) finds in favour of a non-linear relationship between order flow and exchange rates, with the relationship stronger during periods of high bid-ask spreads and volatility, and low traded volumes. However, there is some variation in the significance of results across exchange rates, and for different measures of market conditions, suggesting that Luo's results may be overly sensitive to the sample period under study.

Osler (2002) finds statistically significant evidence in favour of the view that stop orders have a larger price impact during afternoon trading in New York, when market liquidity is traditionally relatively low, than during New York morning trading sessions that overlap the afternoon London trading session. In a similar vein, Froot and Ramadorai (2001) conclude that the relationship between exchange rate returns and order flow depends upon whether returns are permanent or temporary. In particular, order flow exhibits some explanatory power for temporary shifts in exchange rates around Fair Value levels but has no explanatory power for innovations in Fair Value levels. This evidence is consistent with a central hypothesis of market microstructure theory that suggests order flow should have more explanatory power for high frequency exchange rate returns, with macroeconomic fundamentals more relevant over longer-term horizons (Sarno and Taylor, 2002).

In conclusion, the preceding discussion has highlighted the variety of methodologies that researchers have adopted in an effort to assess the validity of a core hypothesis of the market microstructure literature: that order flow is the crucial link in the mechanism by which dispersed private information in the foreign exchange market becomes embedded into prices. This mechanism incorporates two stages, with customers first revealing private information to dealers, who, second, share this information in the process of spreading credit and market risk assumed from customers across the interdealer market.

The results of this research have been inconclusive, often implying a significant contemporaneous correlation between order flow and exchange rate returns but generally indicating little out-of-sample predictive power. These results are typically based upon short spans of data that are often unrepresentative of the data available on a real-time basis to the wider market, particularly customers. As such, they have done little to dispel the scepticism surrounding the market microstructure literature in wider academia and amongst foreign exchange customers (for example, see Rogoff, 2002).

4.4 Data

We examine the relationship between order flow and exchange rate returns using three databases, one of interdealer order flow and the others customer order flow from two major investment banks. Our first data set was kindly provided by Richard Lyons and is identical to the data set used in Evans and Lyons (2002a). It contains eighty daily observations on interdealer order flow for mark-dollar and yen-dollar during the sample period May 1st to August 29th, 1996. The data were collected from Reuters D2000-1 interdealer service and are defined as the difference between the number of buyer-initiated and seller-initiated trades. All exchange rates are defined as the domestic price of foreign currency. We examine these data in order to critically assess the results of Evans and Lyons (2002a), to test the validity of their stated causal relationship from order flow to exchange rates, to assess the predictive value of interdealer order flow, and to test the robustness of their results to the introduction of realistic publication lags.

Our second data set was kindly provided by JPMorgan Chase (JPM). This institution accounts for about 7% of daily customer foreign exchange order flow, equivalent to approximately US$4.6 billion (Euromoney, 2003); as reported in Table 1, this represents the fourth largest market share. This data set comprises JPM's FX Flow Indicator (FXI), which incorporates raw data collated from the bank's global daily book of business on the aggregate daily US dollar amount of net purchases of a particular currency transacted with JPM through its custody business (JPMorgan, 2002). The data span the sample period January 1, 1999 to June 9, 2003, or 1151 daily observations, for euro-dollar, yen-dollar, sterling-dollar and Swiss franc-dollar.

Raw JPM customer flow data is manipulated in three steps to generate the FXI. First, the daily flow into currency i is divided by the average absolute daily net flow into this currency over the past twelve months; this provides some indication of the strength of net daily flows into currency i. Second, this scaled daily flow is multiplied by 100 to generate an index. Third, this index is smoothed using a five-day moving average. This smoothed index is the FXI. A value of +100 for the FXI of currency i indicates net inflows during the past five days equivalent to the average inflow for the last year (JPMorgan, 2002). The FXI is reported with a one day lag[79].

Our third data set is kindly provided by the Royal Bank of Scotland (RBS). This institution accounts for about 3% of daily customer foreign exchange order flow, equivalent to approximately US$2.0 billion (Euromoney, 2003); from Table 1, this is the twelfth largest market share. Our data set comprises RBS's FX Flow Index (FFI), for euro-dollar, yen-dollar, sterling-dollar and Swiss franc-dollar, over the sample period October 1, 2001 to 15 May, 2003. Data for July 2002 is missing (we take account of this gap in our analysis), so that our sample incorporates a total of 387 daily observations. The FFI is defined as a measure of net daily customer purchases of currency i (expressed in terms of the US dollar) transacted with RBS (RBS, 2003). The index is scaled to lie within the values +1 (all buys) and -1 (all sells).

[79] The raw data are manipulated in this manner primarily to ensure customer confidentiality. This applies equally to RBS order flow data below. Although this manipulation will dilute the information content of these data, this is an important finding: preferred customers see only the manipulated data; only dealers of the owning institution see the raw flow on a real-time-but uncollated-basis.

4.5 Empirical Results

4.5.1 The Evans-Lyons data set

In their path-breaking study of foreign exchange market microstructure, Evans and Lyons (2002a) analyse the ability of interdealer order flow data to explain the daily variation of mark-dollar and yen-dollar during the sample period May through August 1996. A visual examination of the data suggests a number of interesting features. First, and consistent with the discussion above on trading strategies, cumulative interdealer order flow appears to be relatively persistent for yen-dollar, with no tendency to revert back towards zero. And although mark-dollar order flow is less persistent, the half-life of deviations from zero still appears inconsistent with the traditional academic assumption that interdealer inventory positions are reduced to zero at the end of every trading day due to the absence of overnight risk budgets. Indeed, it seems to support the description of market structure that we provide in an earlier section of this paper. Second, there is a high contemporaneous correlation between the order flow series and the associated exchange rates, with correlation coefficients of 0.80 for mark-dollar and 0.77 for yen-dollar. At a purely subjective level, therefore, evidence exists to support the hypothesis that interdealer order flow has some explanatory power for the behavior of daily exchange rate returns.

To examine this relationship more rigorously, Evans and Lyons (2002a) estimate the following equation by ordinary least squares (OLS):

$$\Delta s_t = \beta_1 \Delta(i_t - i_t^*) + \beta_2 X_t + \varepsilon_t, \tag{59}$$

where Δs_t is defined as the daily log change in the exchange rate from 4.00 p.m. GMT on day $t-1$ to the same time on day t; $\Delta(i_t - i_t^*)$ is the change over the same period in the one-day domestic-foreign interest differential, which is the authors' proxy for relevant exchange rate fundamentals; X_t is interdealer order flow between 4.00 p.m. GMT on day $t-1$ and the same time on day t; and ε_t is an error term, assumed independently and identically distributed with mean zero and variance σ.

Before discussing Evans and Lyon's results, we note that it appears reasonable to question whether equation (60) is misspecified by inclusion of the change, rather than the level of interest rate differentials (Engel, 1998). The answer depends upon the statistical characteristics of each series. The typical conclusion within studies that focus upon the floating exchange rate period is that nominal exchange rates exhibit a unit root process, or are I(1) series. By contrast, over any reasonable data span, interest differentials between major economies such as Germany, Japan and the US must by definition be stationary series, or I(0). If we accept this consensus–the Evans and Lyons (2002a) data span is too short to allow rigorous testing–then equation (60), by combining variables integrated at different orders, is spurious and without economic meaning. To see this, consider the behaviour of these two series in the wake of a shock: the shock to the exchange rate will be persistent, whereas the interest rate differential will tend to revert back to its mean value relatively quickly. Consequently, the two series will diverge and cannot exhibit long-run co-movement, or cointegration.

Leaving issues of misspecification to one side, on the basis of equation (60), Evans and Lyons (2002a) conclude that interdealer order flow data can explain a significant proportion of contemporaneous daily exchange rate variation. Our reworking of their analysis yields similar results (Table 2)[80]. The R^2 of the estimated equation, assuming no publication lags, is 0.64 for mark-dollar and 0.44 for yen-dollar. This is an impressive result, particularly as it relates to daily returns which, traditionally, are assumed to be distributed as random. Furthermore, the estimated coefficients for explanatory variables in both equations are correctly signed and significant, other than interest differentials for mark-dollar. Thus, an increase in interdealer order flow in the Evans and Lyons framework indicates net US dollar purchases. This is consistent with a depreciation of the exchange rate, implying a positive coefficient on this variable as well, which is borne out by the estimation results. Similarly, to ensure that risk-adjusted returns are equivalent across an exchange rate, uncovered interest parity dictates that higher domestic interest rates should be consistent with depreciation (that is, a rise) in the domestic currency. Consequently, one would expect to observe a positive coefficient for the interest differential term in both equations, and this is also borne out by the estimation results.

Omitting order flow from equation (60) and simply regressing exchange rate returns on interest differentials reduces the R^2 statistic in both cases to below 1%, and renders estimated coefficients insignificant at the 5% level. This leads Evans and Lyons (2002a) to conclude that daily variation in both exchange rates is largely explained by interdealer order flow, with little role at this frequency for fundamental variables. To this end, we are able to confirm the finding, on the basis of a Wald test, that the change in significance in the interest differential between the two specifications is consistent with omitted variable bias in the simplified equation (Evans and Lyons, 2002). Accordingly, on the basis of the evidence presented so far daily interdealer order flow has significant explanatory power for daily exchange rate returns.

As it stands, this is an extremely interesting result and is certainly worthy of further consideration, on three counts; publication lags; Granger-causality; and the accuracy of associated exchange rate forecasts.

Publication lags Any viable investment strategy can incorporate only data publicly available in advance of the introduction of active currency hedges into portfolios. Evans and Lyons (2002a) explain daily exchange rate returns from period $t-1$ to t using contemporaneous order flow and interest rate data. This implies perfect foresight. Correcting for this flaw means introducing a one-day lag into both explanatory variables in equation (60). The estimation results from the revised equation are presented in Table 2. With appropriate publication lags, the relationship between daily exchange rate returns and interdealer order flow becomes insignificant, and the R^2 of both equations falls to zero.

[80] We are unable to account for small differences in coefficient estimates and R^2 statistics compared with Evans and Lyons (2002a).

Granger-causality Evans and Lyons (2002a) assume the existence of an appropriate causal relationship running from interdealer order flow to exchange rate returns, but do not provide explicit test results to support this assumption. Accordingly, we perform rigorous Granger-causality tests and report p-values for tests of joint significance of lags from 1 to 10 in Table 3. The results of these tests suggest that the Evans-Lyons hypothesis is incorrect[81]; in fact, the significant causal relationship runs from exchange rate returns to order flow for both exchange rates, albeit only at a 10% significance level for yen-dollar.

Forecast Accuracy Although contemporaneous explanatory power is certainly interesting, the merits of order flow will be determined principally by its ability to improve upon the generally low out-of-sample forecasting accuracy of traditional fundamental exchange rate models. To date this metric has not been rigorously examined.

In a companion paper to Evans and Lyons (2002a), Evans and Lyons (2001) report RMSFEs of exchange rate forecasts for mark-dollar and yen-dollar over one-day, and one- and two-week horizons derived from a naive random walk and a recursive estimation of Evans and Lyons (2002a). These forecasts take the form:

$$\sum_{j=1}^{k} \Delta s_{t+j} = \beta_1 \sum_{j=1}^{k} \Delta(i_{t+j} - i^*_{t+j}) + \beta_2 \sum_{j=1}^{k} X_{t+j} + \varepsilon_t, \tag{60}$$

for $j = 1, 5, 10$. The forecasts are based upon traditional Meese-Rogoff (1983a, b) evaluation criteria. These criteria involve, in this application, using data over the initial thirty-nine observations of the Evans and Lyons data set to generate initial coefficient estimates of the relationship between exchange rates, interdealer order flow and interest differentials. This estimated relationship is then used to forecast the change in the exchange rate at time $t + j$ based upon observed values of order flow and interest differentials also at time $t + j$. These initial coefficient estimates are then recursively updated, and the forecasting exercise repeated. Consequently, the Meese-Rogoff criteria unrealistically assume perfect foresight with regard to explanatory variables. We replicate this analysis in Table 4. Consistent with the results of Evans and Lyons (2001), we conclude that the use of interdealer order flow under the assumption of perfect foresight generates more accurate forecasts across all horizons and for both exchange rates than a random walk model.

In order to assess the significance of this improvement in forecast accuracy, we calculate the Diebold-Mariano (DM; Diebold and Mariano, 1995) test statistic for equality of forecast accuracy; see the previous chapter for a detailed explanation of the DM statistic. As we report in Table 4, the DM test statistic indicates that the reduction in forecast errors achieved by modelling daily

[81] As the small sample properties of Granger-causality tests is unclear, this result should be interpreted with a degree of caution.

exchange rate returns as a function of interdealer order flow and interest differentials within the Evans and Lyons (2002a) framework in place of a random walk model is not significant, at any of our forecast horizons. This finding represents an important contradiction of a central hypothesis of the market microstructure literature.

The assumption of perfect foresight implicit within the Meese-Rogoff criteria is not appropriate in an investment portfolio context. Accordingly, we repeat the above analysis (equation (61)) using only data available at the date exchange rate forecasts are compiled, that is

$$\sum_{j=1}^{k} \Delta s_{t+j} = \beta_1 \sum_{j=1}^{k} \Delta(i_{t+j-1} - i^*_{t+j-1}) + \beta_2 \sum_{j=1}^{k} X_{t+j-1} + \varepsilon_{t+j}, \tag{61}$$

where $j = 1, 5, 10$. We term these limited information forecasts, and report the results in Table 5. In this case, and at all horizons, RMSFE statistics associated with the random walk forecasts are lower for both exchange rates than those generated by the Evans and Lyons model. Furthermore, DM test statistics indicate that this reduction in RMSFEs in favour of the random walk model is significant, other than at a one-day horizon for yen-dollar. Under realistic trading conditions, the Evans and Lyons (2002a) model generates exchange rate forecasts that are inferior to a naive random walk.

We also compute a series of long horizon forecasts on the basis of Evans and Lyons (2002a). Long horizon forecasts take the form,

$$s_{t+k} - s_t = \beta_{1k} \Delta(i_t - i^*_t) + \beta_{2k} X_t + \varepsilon_t, \tag{62}$$

for $k = 1, 2, ..., 10$. Table 6 reports the p-values associated with estimated coefficients at forecast horizons from one to ten days. These p-values are calculated using a non-parametric bootstrap that involves 5000 simulations of a naive random walk model for both exchange rates over every forecast horizon. This bootstrap defines the shape of the associated probability density function for each estimated coefficient. As such, it provides a robust measure of the significance of coefficient estimates derived from the Evans and Lyons (2002a) model. Any p-value below 0.05 implies a rejection of the random walk null hypothesis in favour of the Evans and Lyons model at the 5% significance level.

Consistent with the limited information forecasts above, the results of long horizon forecasts indicate that interdealer order flow has no significant predictive power for daily exchange rates returns at any horizon. Interest differentials are also generally insignificant, although at intermediate horizons for mark-dollar some evidence of weak significance for this variable is apparent. Again, therefore, these results represent an important contradiction of a central hypothesis of the microstructure literature.

To conclude this section, Evans and Lyons (2002a) present an excellent in-sample explanation of daily exchange rate variation that emphasises the importance of order flow over fundamental variables for data samples that incorporate high frequency observations. Their findings represent an important step in improving our understanding of the transmission mechanism by which dispersed private information is embedded into prices in the foreign exchange market. Disappointingly–and yet consistent with the general experience of the past three decades of empirical exchange rate modelling–the predictive power of interdealer order flow data, albeit on the basis of a very short time span, is poor and implies that these data, sampled on a daily frequency, cannot be used to improve the quality of exchange rate forecasting or portfolio investment decision-making.

4.5.2 JPMorgan Chase (JPM) FXI Aggregate Customer Order Data

It is an open question whether predictive performance can be improved using alternative sources of order flow, and particularly customer data. Intuitively, this approach should lead to some improvement as customer orders arguably represent the main source of private information in the foreign exchange market. We address this issue in this section using customer order flow from JPM, and in the following section from RBS.

Correlation analysis suggests a looser relationship between JPM customer order flow data and exchange rates than is the case for the interdealer order flow of Evans and Lyons (2002a). Whole-sample coefficients are generally below 0.20 in absolute terms for all four exchange rates (Table 7). The sign of correlations is generally consistent with theoretical priors, with net purchases of the domestic currency (US dollar) consistent with an appreciation (depreciation) of the exchange rate. An exception is sterling-dollar in 1999 when net flows into both currencies are consistent with a depreciation of the exchange rate.

Our next step is to re-estimate the Evans and Lyons (2002a) model (equation (60) above) using the JPM FXI database for each of the four exchange rates. The results of this exercise, with and without appropriate publication lags[82], are reported in Table 8. As the FXI is defined as net inflows into currency i from all other currencies we have included separately in equation (4) both net inflows into the domestic currency and into the dollar; as the predominant currency within bond and equity indices, the dollar is included as a proxy for flows into all other currencies.

The most striking contrast between the results reported in Tables 8 and 3 is the magnitude of estimated R^2 statistics. Using JPM data, and assuming no publication lags, a maximum of just 2% of daily exchange rate variation is explained by the Evans and Lyons model. This is far worse than the results of estimated equations that use interdealer order flow.

The magnitude of absolute coefficient estimates for order flow reported in Table 8 do not lend themselves to economic interpretation given the extent of manipulation of the underlying raw data discussed above. But the sign and relative magnitude of estimated coefficients are instructive: coefficient signs on both order flow terms are consistent with a priori expectations for all four exchange rates, with net purchases of domestic currency (dollars) consistent

[82] These data are published with a one day lag (i.e. t+1).

with an appreciation (depreciation) of the exchange rate[83]. Coefficients are statistically significant for net purchases of euros and Swiss francs, and for US dollar purchases for yen-dollar, sterling-dollar and Swiss franc-dollar. In addition, the magnitude of estimated coefficients indicates that the strongest contemporaneous relationship between net purchases of domestic currency and exchange rate returns exists for Swiss franc-dollar.

Estimated coefficients for the interest differential term are also consistent with a priori expectations, except for sterling, where an increase in domestic interest rates relative to US rates is found to be consistent with an appreciation of the domestic currency.

Publication Lags Consistent with our reworking of Evans and Lyons (2002a), explanatory power falls to zero for all exchange rates once appropriate publication lags have been incorporated into equation (60). Furthermore, coefficient estimates in Table 8 are now insignificantly different from zero, except for interest differentials which are weakly significant for sterling-dollar (but still incorrectly signed). In addition, inclusion of a one day publication lag causes the sign of many coefficients to switch.

Granger-causality Granger-causality tests between JPM FXI customer order flow data and exchange rate returns also offer little reason for optimism (Table 9). Contrary to theoretical prediction, causality runs strictly from exchange rate returns to customer order flow for euro-dollar, sterling-dollar and Swiss franc-dollar. For yen-dollar, there is evidence of two-way causality between returns and net purchases of the domestic currency; for net dollar purchases, the strict, perverse causal relationship observed for other exchange rates is evident also for this exchange rate.

The general presence of a perverse causal relationship between order flow and exchange rate returns could reflect a number of factors. First, it may simply be that aggregate customer order flow has no predictive value for exchange rate returns. This conclusion, which we examine in more detail below, is consistent with the view of Andersen, Bollerslev and Das (1998), who suggest that market microstructure theories are typically not designed to provide quantitative predictions, but merely a qualitative characterisation of the pattern that is likely to arise in some market variables, including exchange rate returns.

Second, JPM FXI order flow data may simply be unrepresentative of market trends, regardless of publication lags. This may in turn be due either to a relatively small absolute market share for JPM within total customer order flows, or to pre-filtering of the raw order flow data that greatly diminishes their information content. Both explanations are certainly possible. Although JPM boasts the fourth largest market share in the customer order space of the foreign exchange market, it transacts only 7% of total daily turnover (Euromoney,

[83]Coefficient signs on the order flow variable in the no-lag variant are reversed compared with the Evans and Lyons results in Table 3. This reflects a difference in the construction of these data between the different sources: in the Evans and Lyons data a positive order represents a purchase of dollars; in this case it represents a sale.

2003). It is an impossible task, however, to aggregate customer order flow data across transacting banks to achieve a database that unequivocally covers a critical mass market share, for a number of reasons. Crucially, most investors do not have access to two or more order flow data sets, and many do not enjoy access even to one. For those that do have access to two or more, differences in data measurement and aggregation make the task of compiling a composite database impractical. Similarly, it is an unavoidable fact that practitioners outside of the owning institution gain access only to pre-filtered, indexed order flow data. This filtering, necessary to ensure customer anonymity, must dilute the information content of the data, potentially to the point of rendering them practically useless as inputs into exchange rate forecasting models.

Forecast Accuracy Table 10 reports the RMSFEs of forecasts prepared under Meese-Rogoff (1983a, b) perfect foresight criteria using JPM customer order flow data within both Evans and Lyons and random walk models. As above, forecasts are calibrated over one-day, and one and two-week horizons. The final column in Table 10 reports the associated DM test statistic, with a negative value indicative of an improvement in forecast accuracy moving from the random walk to the Evans and Lyons model.

The results of this exercise indicate that perfect foresight forecast errors generated by the Evans and Lyons model incorporating JPM FXI customer order flow data are generally smaller than those generated by the random walk model. In addition, DM test statistics indicate that this improvement is statistically significant in many cases. This is a marked improvement on the results derived from interdealer order flow and offers at least some solace to the microstructure literature.

This improvement in forecast accuracy does not transfer to out-of-sample analysis. As Table 11 indicates, for euro-dollar, yen-dollar and Swiss franc-dollar, the random walk model now generates a statistically significant reduction in forecast errors compared with the Evans and Lyons model. For sterling-dollar, forecast errors generated by the random walk model are consistently smaller than the Evans and Lyons model, but this improvement is only significant at a two-week horizon.

In Table 12 we report p-values for long horizon forecasts from one to ten days ahead For euro-dollar and Swiss franc-dollar, these forecasts fully confirm the results of the preceding analysis: customer order flow provides no improvement in forecast accuracy relative to a naive random walk. Indeed, the only significant coefficient at the 5% level for either of these exchange rates is euro-dollar interest differentials for a four-day-ahead horizon; we are inclined not to put too much emphasis upon this result, given its relative isolation. In addition, reported F-statistics for both exchange rates indicate that coefficients are jointly insignificant at each forecast horizon.

For yen-dollar and sterling-dollar, the results are a little more favourable to the market microstructure literature. For yen-dollar, net purchases of domestic currency have a significant explanatory role for exchange rate returns from six to ten days ahead. This result may be consistent with the finding above of two-way causality between exchange rate returns and net purchases of yen, with particular customer groups reacting to an earlier move in yen-dollar by subsequently adjusting their positions in this exchange rate. Most logically, this argument should apply to corporations who adjust their hedging behaviour depending upon the level of the exchange rate, and technical and model-based investors who invest on the basis of directional trend analysis. Taken together, all three coefficients for yen-dollar are jointly insignificant at every horizon.

For sterling-dollar, results are perplexing. Net purchases of sterling exhibit no significant explanatory power for exchange rate returns at a 5% level, although for horizons beyond six days there is some evidence of weak significance. By contrast, net purchases of US dollars, our proxy for order flow into all other currencies, is significant at all forecast horizons except ten days ahead. The interest rate term is also significant at a few horizons. In addition, at intermediate horizons estimated coefficients are jointly significant. It is not clear why net customer order flow into other currencies would be significant for the behaviour of future sterling-dollar exchange rate returns when net purchases of sterling are not. Given all of the accumulated evidence above that aggregate JPM customer order flow generally is not a significant explanatory variable of sterling-dollar returns, we are inclined to caution against over-interpretation of this result. Analysis of RBS customer order flow will allow us to consider the validity of this result further, and it is to this analysis that we now turn.

4.5.3 Royal Bank of Scotland (RBS) FFI Aggregate Customer Order Data

We repeat the analysis of customer order flow, for the same four exchange rates, using aggregate customer data from RBS. The RBS data set comprises 387 daily observations from October 1, 2001 to 15 May, 2003. For reasons unknown, data for July 2002 were not collected by RBS. This discontinuity clearly shortens the usable data sample for many of the procedures we have discussed above. As a result, for each of the forecasting exercises we separate the data into two halves, using data for October 2001 through June 2002 to generate initial recursive coefficient estimates and the data for August 2002 through May 2003 to generate in-sample, out-of-sample and long horizon forecasts. For the regression analysis, we surmount this discontinuity by including dummy variables as a naive splice.

Correlation analysis reported in Table 13 provides a mixed picture. The sign of correlation coefficients is only consistent with theoretical priors across the full data sample for sterling-dollar, with net order flow consistent with an appreciation of sterling. For euro-dollar, the sign of correlations switches between the two halves of the sample; for yen-dollar and Swiss franc-dollar the correlation of aggregate order flow and exchange rate returns is perverse.

The results of the reworking of the Evans and Lyons model (equation (60)) are also patchy (Table 14). Estimated order flow coefficients are correctly signed but not significant at a 5% level for euro-dollar and sterling-dollar[84]. For yen-

[84] As in the previous section, prior manipulation of the raw order flow data to create the

dollar and Swiss franc-dollar, the converse is true: coefficients are significant, but wrongly signed. R^2 statistics are generally close to zero; the sterling-dollar equation achieving the highest explanatory power, at 6%. Estimated coefficients for the interest rate term are generally consistent with previous academic findings that contradict uncovered interest parity. These results are much worse than the results of Evans and Lyons (2002a), and no better than a traditional fundamental-based model of daily exchange rate variation.

Publication Lags Once we remove perfect foresight, order flow coefficients become insignificant for all exchange rates, and explanatory power remains poor. Evidence of coefficient sign switching between exchange rate equations is consistent with the findings of Evans and Lyons (2002b) and Lyons (2003) and suggests that the underlying relationship exploited by the Evans and Lyons (2002a) model is unstable.

Granger-causality Granger-causality tests on aggregate FFI data are reported in Table 15. They indicate a significant causal relationship for yen-dollar that runs from the exchange rate to customer order flow. A similar, albeit weakly significant, result is evident for the Swiss franc-dollar. For none of our four exchange rates does FFI customer order flow cause exchange rate returns. These findings are largely consistent with the results from interdealer and JPM aggregate customer order flow data.

Forecast Accuracy Table 16 reports the results of one-day, and one and two-week forecasts under Meese-Rogoff (1983a, b) criteria. For yen-dollar, DM test statistics indicate a significant improvement in forecast accuracy moving from the naive random walk to the Evans and Lyons model incorporating aggregate customer order flow at both one and two week horizons, and an insignificant improvement at a one-day horizon. For euro-dollar and Swiss franc-dollar this-insignificant-improvement is also evident for one- and two-week horizons. By contrast, for sterling-dollar the random walk model generates smaller forecast errors than the Evans and Lyons model.

When we relax the assumption of perfect foresight (Table 17), results generally look very different. For euro-dollar, the Evans and Lyons models now achieves an insignificant improvement in forecast accuracy relative to the random walk model at a one day horizon. For all other horizons and for all exchange rates, the random walk model generates smaller forecast errors than Evans and Lyons. In the case of yen-dollar and sterling-dollar, this improvement in favour of the random walk model is significant for one and two week ahead forecasts.

Moving to the long horizon forecasts (Table 18), reported p-values provide scant evidence in favour of the order flow literature. Excluding isolated evidence of a (weakly) significant relationship at a two day horizon for euro-dollar, and four and ten days for yen-dollar, these results are generally consistent with the preceding findings from interdealer and JPM customer order flow. Although

FFI index inhibits any economic interpretation of the estimated order flow coefficients.

order flow intuitively represents the mechanism by which disbursed private information becomes embedded within exchange rates, knowledge of this mechanism and real-time access to available sources of interdealer or customer order flow data-filtered and indexed to maintain customer confidentiality-appears to be of no practical value to either forecasters or investment portfolio managers.

4.5.4 Royal Bank of Scotland (RBS) FFI Disaggregated Customer Order Data

Beyond the issues of data manipulation and market share that we have already discussed, our lack of success in generating results generally supportive of the core hypotheses of the market microstructure literature may reflect our concentration so far upon aggregate customer order flow. Marked heterogeneities exist within the customer segment of the foreign exchange market that imply differences in the way various customers react to news (Lyons, 2003). Intuitively, a currency overlay manager should exhibit a different trading pattern to a CTA, and a corporation's hedging activity will be different to the foreign exchange activity of a central bank. This observation suggests that examination of customer order flow disaggregated by customer type may uncover more supportive evidence for the role of market microstructure in price determination.

Accordingly, our final data set, also kindly provided by RBS, separates customer order flow into four distinct customer groupings: corporations, real money managers, leveraged money managers and other customers. This final category includes central banks, non-leveraged system accounts and non-reciprocal banks[85]. Clearly, even our disaggregated data is, for reasons of client confidentiality, pre-filtered and manipulated into individual indices for each of the four subgroups. Furthermore, the allocation of RBS clients into these categories will be arbitrary to a degree, with some participants spanning more than one bucket, for instance a number of currency overlay managers also offer clients leveraged currency products. Nonetheless, these data represent a unique opportunity to quantitatively test the core hypotheses of the market microstructure literature based on order flow data from a major investment bank. Although other studies have examined disaggregated customer order flow, this analysis has either been qualitative (Fan and Lyons, 2001) or based upon order flow data provided by a non-reciprocal bank in a minor exchange rate (Bjønnes, Rime and Solheim, 2003). To the best of our knowledge, this is the first study that has quantitatively assessed the value of disaggregated customer order flow available to the wider market from a major investment bank on a real-time basis across a number of the most liquid exchange rates.

Correlation analysis reported in Table 19 highlights important differences in the behaviour of participant groups. Correlations between other customer order flow and exchange rate returns are consistent with a priori expectations, except for Swiss franc-dollar, as well as yen-dollar in the first half of the sample. For

[85] Non-reciprocal banks are defined as smaller banks that make prices in their local markets but that are price-takers in the wider foreign exchange market, outsourcing their liquidity requirements in major exchange rates to larger investment banks, such as JPM or RBS.

real and leveraged customer flows the correlation with exchange rate returns is more equivocal, and for corporate customers the correlation is reversed, for all exchange rates: in this case, net purchases of the domestic currency are consistent with a depreciation, suggesting that corporations react to, rather than pre-empt price innovations.

Publication Lags As in previous sections, we now re-estimate the model of Evans and Lyons (2002a) under the assumption of perfect foresight, this time replacing interdealer order flow in the original specification with our four disaggregated customer order flow series. This exercise provides a number of interesting results (Table 20). First, R^2 statistics remain generally close to zero, except for sterling-dollar, where 10% of daily variation is explained by this variant of the Evans and Lyons model. This is much lower than the results of Evans and Lyons (2002a) using interdealer order flow, but is an improvement upon results generated using aggregate customer order flow; in addition, once appropriate publication lags have been included the explanatory power of the sterling-dollar equation falls back to 5%[86].

Second, although the magnitude of estimated coefficients again cannot readily be interpreted their sign and significance are instructive. A number of the coefficients are significant, bearing out the predictions of the microstructure literature. But this result is greatly impacted by the inclusion of publication lags, and there is also evidence of parameter instability and sign switching between the perfect foresight and lagged versions of the model that cautions against overinterpreting these results. Interestingly, there are no significant coefficients for euro-dollar regardless of publication lags.

Granger-causality For all of the exchange rates Granger-causality tests reveal no significant causal relationships running from disaggregated order flow to exchange rate returns that prevail over both halves of the sample data (Table 21)[87]. By contrast, net yen purchases by corporations and leveraged money managers are Granger-caused by returns over both halves; this last result is consistent with the findings of Fan and Lyons (2001) who conclude that this customer group provided liquidity to the market at the time of the substantial appreciation of yen-dollar during October 1998. Granger-causality running from exchange rate returns to leveraged money manager order flow is also apparent for Swiss franc-dollar, and the same result is evident for sterling-dollar corporate order flow. This perverse causal relationship for leveraged money manager order flow may reflect the predominance at short horizons of technical, trend-following investors whose investment style by design is reactive to price innovations, compared with fundamental-based flows of active managers that should incorporate a predictive element. A similar argument applies to the hedging activity of corporations, as suggested also by the correlation analysis above.

[86]These data are published with a one day lag, i.e. t+1.

[87]As discussed above, we separate the data into two halves due to missing data in July 2002.

Forecast Accuracy The results of Meese-Rogoff forecasts using disaggregated customer order flow are generally consistent with the findings from JPM and RBS aggregate customer order flow data (Table 22). For every forecast horizon and exchange rate, except euro-dollar at one day, the Evans and Lyons model achieves an improvement in forecasting accuracy relative to a naive random walk. DM test statistics indicate that these improvements are more widely significant than is the case for either of the aggregate databases discussed above. Leaving to one side issues of perfect foresight, these results would suggest that disaggregated customer order flow data has a greater ability to predict in-sample exchange rate returns than either daily interdealer or aggregate customer order flow.

Although this result supports a central assumption of the microstructure literature, it depends critically upon the assumption of perfect foresight. In Table 23, we relax this assumption. As indicated by the DM test statistic, the random walk model now generates more accurate forecasts than the Evans and Lyons model, except for Swiss franc-dollar at a one-day horizon. In addition, this improvement is statistically significant in many cases. Moreover, from Table 24 the results of long horizon forecasts provide scant evidence of significant forecasting power. Indeed, the only real pocket of information exists for net yen purchases by real money managers on a three to five-day horizon, and f-tests indicate joint insignificance of coefficients for all exchange rates at every forecast horizon.

In the light of these forecasting results, as well as persistent evidence of the low explanatory power and perverse Granger-causality throughout much of our analysis of both aggregate and disaggregated customer order flow data, it is difficult to avoid the conclusion that order flow, in the form available to the majority of practitioners, has little or no ability to improve upon the out-of-sample performance of fundamentals-based exchange rate models.

4.6 Conclusion

Traditional models of exchange rate determination concentrate upon the relationship between exchange rate movements and innovations in economic fundamentals. Their ability to explain in-sample exchange rate returns has been persistently low, although some improvement has been achieved with the application of non-linear modelling techniques that recognise the speed of mean reversion of spot exchange rates to equilibrium values depends crucially upon the size of misalignment (Taylor, Peel and Sarno, 2001). But non-linear modelling has generally not led to a commensurate improvement in the out-of-sample predictive ability of fundamental exchange rate models and this remains generally poor, particularly in the context of point forecasting exercises (although see Killian and Taylor, 2003 and Clarida, Sarno, Taylor and Valente, 2003).

In an effort to improve upon this generally poor track record, much recent research has focused upon market microstructure. Proponents of this approach typically argue that one measure in particular–order flow–may represent the missing link in the process by which dispersed private information is embedded within exchange rates. This process has two stages. First, customers initiate orders with dealers in response to private information for instance relating to the release of public macroeconomic information or to changes in investor risk-return preferences. Second, dealers spread the risk assumed from customers during these trades across the interbank market.

The microstructure literature draws support and scepticism in equal measure. Few disagree with the central hypothesis that order flow is the mechanism by which private information is embedded in exchange rates. Much more disharmony surrounds the assessment of the practical value of this hypothesis. In this paper we set out to provide a rigorous investigation of the relationship between order flow and exchange rate returns, using both interdealer and customer order flow data. In addition, we separated customer order flow into data aggregated across all customers and disaggregated by customer group.

We conclude that the ability of data available to the wider market on a real-time basis to improve upon the forecasting accuracy of fundamental-based models is generally weak. In addition, and in contradiction with theoretical priors, we find widespread evidence of a strict Granger-causal relationship that runs from exchange rate returns to customer order flow. This result is consistent with evidence presented by Payne and Vitale (2002), DPL (2002) and Froot and Ramadorai (2001). We discuss a number of factors that may explain our results. These include market share issues of sampled databases, pre-filtering and indexation of data, but also the validity of hypotheses that lie at the core of the microstructure literature. No single explanation can provide a complete answer. But as our study employs customer order flow from two major investment banks as well as interdealer order flow, for a range of exchange rates and sample periods, it seems reasonable to conclude that our results are relatively robust.

It should also be stressed that for those participants–typically dealers at major investment banks–who are able to sample interdealer and customer order flow at very high frequencies, including on a tick-by-tick basis, and in a raw, unmanipulated form, these data may represent an important, and profitable, source of private information (Bjønnes and Rime, 2001b). We are not in a position to test this possibility, and instead leave this issue on the agenda for future research.

In addition, our results do not invalidate the hypothesis that private information and persistent profit opportunities coexist in the foreign exchange market. Indeed, performance data from the currency overlay industry indicate that they do (Baldridge, Meath and Myers, 2000; Hersey and Minnick, 2000). Our results also do not invalidate other aspects of the microstructure literature, and particularly intra-day volatility studies that have achieved a demonstrable ability to predict and practically exploit significant volatility shifts associated with macroeconomic policy announcement events (Melvin, Sager and Taylor, 2003). But the results presented in this paper do suggest that outside of a few, particularly well informed investors who observe order flow data on an unfiltered, tick-by-tick basis, knowledge of customer or interdealer order flow cannot help improve the quality of exchange rate forecasting or the profitability of investment portfolio decision-making. From this perspective, we have confirmed the results of Meese and Rogoff (1983a, b): exchange rate forecasting remains a hazardous occupation even when the forecaster is equipped with order flow data.

References

[1] Anderson, T.G., T. Bollerslev and A. Das (1998), Testing for Market Microstructure Effects in Intraday Volatility: A Reassessment of the Tokyo FX Experiment. NBER Working Paper # 6666.

[2] Anderson, T.G., T. Bollerslev, F.X. Diebold and C. Vega (2002), Micro Effects of Macro Announcements: Real-Time Price Discovery in Foreign Exchange. NBER Working Paper # 8959.

[3] Baird, D. H. (2002), Integrating eFX with Institutional Investor Needs. JP-Morgan Fleming Asset Management Mimeo.

[4] Baldridge, J., B. Meath, and H. Myers (2000), Capturing Alpha Through Active Currency Overlay, Frank Russell Company.

[5] Bank for International Settlements (2002), Triennial Central Bank Survey of Foreign Exchange and Derivatives Market Activity 2001. Final Results. www.bis.org.

[6] Bank for International Settlements (2004), Triennial Central Bank Survey of Foreign Exchange and Derivatives Market Activity 2004. Preliminary Results. www.bis.org.

[7] Bates, R.G., M.A.H. Demptster and Y.S. Romahi (2003), Evolutionary Reinforcement Learning in FX Order Book and Order Flow Analysis. Centre for Financial Research, Judge Institute of Management, University of Cambridge Mimeo.

[8] Bjønnes, G.H. and D. Rime (2001a), US Exchange Rates and Currency Flows. Unpublished Mimeo Stockholm Institute for Financial Research. www.sifr.org.

[9] Bjønnes, G.H. and D. Rime (2001b), Customer Trading and Information in Foreign Exchange Markets. Working paper # 4. Stockholm Institute for Financial Research.

[10] Bjønnes, G.H. and D. Rime (2003), Dealer Behavior and Trading Systems in Foreign Exchange Markets. Norges Bank working paper 2003/10. http://www.norges-bank.no/publikasjoner/arbeidsnotater/pdf/arb-2003-10.pdf.

[11] Bjønnes, G.H., D. Rime and H. O. A. Solheim (2003), Volume and Volatility in the FX-Market: Does it Matter Who You Are? Forthcoming in Exchange Rate Modelling: Where Do We Stand? P. De Grauwe (ed.). MIT Press.

[12] Brown, B.W. and S. Maital (1981), What Do Economists Know? An Empirical Study of Experts' Expectations.. Econometrica, 49, 2, 491-504.

[13] Cai, J., Y.-L. Cheung. R.S.K. Lee and M. Melvin (2001), 'Once-in-a-Generation' Yen Volatility in 1998: Fundamentals, Intervention and Order Flow. Journal of International Money and Finance, 20, pp.327-347.

[14] Cheung, Y.-W. and M. D. Chinn (1999), Currency Traders and Exchange Rate Dynamics: A Survey of the U.S. Market". Journal of International Money and Finance 20, 439-471.

[15] Cheung, Y.-W. and C. Y.-P. Wong (2000), A Survey of Market Practitioners' Views on Exchange Rate Dynamics. Journal of International Economics, 51, 401-419.

[16] Citibank (2003), Order Book and Option Vols. Report. Citibank.

[17] Clarida, R.H., L. Sarno, M.P. Taylor and G. Valente (2003), The Out-of-Sample Success of Term Structure Models as Exchange Rate Predictors: A Step Beyond. Journal of International Economics, 60, 61-83.

[18] Consensus Economics (2004), Foreign Exchange Consensus Forecasts. http://www.consensuseconomics.com/.

[19] Covrig, V. and M. Melvin (2001), Asymmetric Information and Price Discovery in the FX Market: Does Tokyo Know More about the Yen? Journal of Empirical Finance, 9, pp.271-285.

[20] Dalio, R. (2002), Inefficiencies and Opportunities in the Currency Markets, in A Guide to Currency Overlay Management, D. Knott (ed.) Deutsche Bank. Global Market Research.

[21] Daníelsson, J. and R. Payne (2001), Measuring and Explaining Liquidity on an Electronic Limit Order Book: Evidence from Reuters D2000-2. LSE Financial Markets Group Mimeo.

[22] Daníelsson, J., R. Payne and J. Luo (2002), Exchange Rate Determination and Inter-Market Order Flow Effects. LSE Financial Markets Group Mimeo.

[23] Diebold, F.X. and R. Mariano (1995), Comparing Predictive Accuracy. Journal of Business and Economic Statistics, 13, 253-62.

[24] Dominguez, K. M. (1999), The Market Microstructure of Central Bank Intervention. NBER Working Paper # 7337.

[25] Engel, C. (1998), Notes on Cointegration of Real Interest Rates and Real Exchange Rates. Unpublished University of Wisconsin mimeo.

[26] Euromoney (2002), 2nd International Forex Forum. www.euromoney.com.

[27] Euromoney (2003), FX Poll 2003. www.euromoney.com/fxpoll2003/fxpollresults.html

[28] European Central Bank (2003), Review of the Foreign Exchange Market Structure. www ecb.int.

[29] Evans, M. and R.K. Lyons (2001), Order Flow and Exchange Rate Dynamics. BIS Papers #2. http://www.bis.org/publ/bispap02.htm#pgtop.

[30] Evans, M. and R.K. Lyons (2002a), Order Flow and Exchange Rate Dynamics. Journal of Political Economy, February, 170-180.

[31] Evans, M. and R.K. Lyons (2002b), Informational Integration and FX Trading. Journal of International Money and Finance, 21, 807-31.

[32] Evans, M. and R.K. Lyons (2003), New Micro Exchange Rate Economics, typescript, U.C. Berkeley.

[33] Fan, M. and R.K. Lyons (2001), Customer Trades and Extreme Events in Foreign Exchange. U.C. Berkeley Mimeo.

[34] Federal Reserve Bank of New York (2003). http://www.ny.frb.org/.

[35] Fisher, P. and R. Hillman (2003), Comments on R.K. Lyons, Explaining and Forecasting Exchange Rates with Order Flows. In Economic Policy Forum: Explaining and Forecasting Exchange Rates With Order Flows. http://www.cepii.fr/anglaisgraph/publications/economieinter/rev96/rev96fischer.pdf.

[36] Flood, R.P. and M. P. Taylor (1996), Exchange Rate Economics: What's Wrong with the Conventional Macro Approach?, in The Microstructure of Foreign Exchange Markets, J A. Frankel, G. Galli and A. Giovannini (eds.), National Bureau of Economic Research. Chicago: University of Chicago Press.

[37] Frankel, J. A. and K. Froot (1985), Using Survey Data to Test Some Standard Propositions Regarding Exchange Rate Expectations. NBER Working Paper # W1672.

[38] Frankel, J.A. and A.K. Rose (1995), Empirical Research on Nominal Exchange Rates, in G.M. Grossman and K. Rogoff (eds.) Handbook of International Economics, III, pp.1689-729. Amsterdam: North-Holland.

[39] Froot, K.A. and K. Rogoff (1995), Perspectives on PPP and Long-Run Real Exchange Rates, in G.M. Grossman and K. Rogoff (eds.) Handbook of International Economics, III, pp.1647-88. Amsterdam: North-Holland.

[40] Froot, K. A. and J.A. Frankel (1989), Interpreting Tests of Forward Discount Bias Using Survey Data on Exchange Rate Expectations. NBER Working Paper # W1963.

[41] Froot, K.A. and T. Ramadorai (2001), Currency Returns, Institutional Investor Flows and Exchange-Rate Fundamentals. State Street Associates Working Paper # 5.

[42] FXAll (2003), FXAll Press Releases. http://www.fxall.com/about/pressrelease46.html.

[43] Gehrig, T. and L. Menkhoff (2001), The Use of Flow Analysis in Foreign Exchange: Exploratory Evidence. Forthcoming, Journal of International Money and Finance.

[44] HSBC (2003), FX Flow Dynamics. April. HSBC.

[45] Ito, T., R.K. Lyons and M.T. Melvin (1998), Is There Private Information in the FX Market? The Tokyo Experiment. The Journal of Finance, June, 1111-1130.

[46] Jansen, D.J. and J. de Haan (2003), Statements of ECB Officials and their Effect on the Level and Volatility of the Euro-Dollar Exchange Rate. CESifo Working Paper # 927.

[47] JPMorgan (2002), Cross-Border Flow of Funds: a Primer. JPMorgan Global Foreign Exchange Research. http://morganmarkets.jpmorgan.com.

[48] Killian, L. and M.P. Taylor (2003), Why is it so Difficult to Beat the Random Walk Forecast of Exchange Rates? Journal of International Economics, 60, 85-107.

[49] Killeen, W.P., R.K. Lyons and M.J. Moore (2001), Fixed Versus Flexible: Lessons from EMS Order Flow. NBER Working Paper # 8491.

[50] Knott, D. (ed.) (2002), A Guide to Currency Overlay Management. Deutsche Bank. Global Market Research.

[51] Kyle, A.S. (1985), Continuous Auctions and Insider Trading. Econometrica, Vol. 53, 6, 1315-1336.

[52] Luo, J. (2001), Market Conditions, Order Flow and Exchange Rates (sic.) Determination. London School of Economics mimeo.

[53] Layard-Liesching, R. (2002), Adding Value Through Currency Management. In A Guide to Currency Overlay Management, D. Knott (ed.) Deutsche Bank. Global Market Research.

[54] Lyons, R. K. (1995), Tests of Microstructural Hypotheses in the Foreign Exchange Market. Journal of Financial Economics, 39, 321-351.

[55] Lyons, R. K. (1996), Foreign Exchange Volume: Sounds or Fury Signifying Nothing? In J. Frankel et. al., The Microstructure of Foreign Exchange Markets. Chicago: University of Chicago Press.

[56] Lyons, R.K. (2001), The Microstructure Approach to Exchange Rates. MIT Press.

[57] Lyons, R.K. (2002a), Foreign Exchange: Macro Puzzles, Micro Tools. Federal Reserve Bank of San Francisco Annual Review.

[58] Lyons, R.K. (2002b), Theoretical Perspective on Euro Liquidity. Economic Policy. Vol. 17, 35, 572-597.

[59] Lyons, R.K. (2003), Explaining and Forecasting Exchange Rates with Order Flows. 3^{RD} Bank of France Economic Forum. http://www.banque-france.fr/gb/fondatio/telechar/R_k_lyons.pdf.

[60] Meese, R.A. and K. Rogoff (1983a), Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample? Journal of International Economics, 14, 3-24.

[61] Meese, R.A. and K. Rogoff (1983b), The Out-of-Sample Failure of Empirical Exchange Rate Models: Sampling Error or Misspecification? in Frenkel, J.A. (ed.), Exchange Rate and International Economics. Chicago: University of Chicago Press.

[62] Melvin, M., M. J. Sager and M.P. Taylor (2003), Can an Old Lady Keep a Secret? A Microstructural study of Monetary Policy Committee Announcements and the Foreign Exchange Market. Unpublished University of Warwick/Arizona mimeo.

[63] Melvin, M. and L. Wen (2003), The Choice of Direct Dealing or Electronic Brokerage in Foreign Exchange Trading. Unpublished Arizona State University Mimeo.

[64] Mende, A. and L. Menkhoff (2003), Different Counterparties, Different Foreign Exchange Trading? The Perspective of a Median Bank. Unpublished University of Hannover Mimeo.

[65] Merrill Lynch (2003), Global FX Paper #3, July 22.

[66] Newey, W. and K. West (1987). "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," Econometrica, 55, 703-708.

[67] NYSE (2003), New York Stock Exchange. http://www.nyse.com/home.html.

[68] Osler, C.L. (2002), Stop-Loss Orders and Price Cascades in Currency Markets. Federal Reserve Bank of New York Staff Report # 150.

[69] Osler, C.L. (2003), Currency Orders and Exchange Rate Dynamics: An Explanation for the Predictive Success of Technical Analysis. Journal of Finance, 58, 5, pp.1791-1820.

[70] Payne, R. (2000), Informed Trade in Spot Foreign Exchange Markets: An Empirical Investigation, Mimeo, Financial Markets Group, London School of Economics.

[71] Payne, R. and P. Vitale (2002), A Transaction Level Study of Central Bank Intervention on Exchange Rates. Unpublished LSE Financial Markets Group Mimeo.

[72] Portes, R. (2002), Comment on Hau, H., W. Killeen and M. Moore, How has the Euro Changed the Foreign Exchange Market? Economic Policy, vol. 17, 34, 149-191.

[73] RBS (2003), FX Flows: Analysis of Customer Flows in the Spot FX Market. RBS.

[74] Rime, D. (2001), Private or Public Information in Foreign Exchange Markets? An Empirical Analysis. www.sifr.org.

[75] Rogoff, K. (2002), Discussion, in Theoretical Perspective on Euro Liquidity, R. K. Lyons, Economic Policy, Vol. 17, 35, 572-597.

[76] Sarno, L. and M. P. Taylor (2002), The Economics of Exchange Rates. Cambridge: Cambridge University Press.

[77] Taylor, M.P. (1995), The Economics of Exchange Rates. Journal of Economic Literature,33, pp.13-47.

[78] Taylor, M.P. (1988), What do Investment Managers Know? An Empirical Study of Practitioner's Predictions. Economica, 55, 218, 185-202.

[79] Taylor, M.P. and H. Allen (1992), The Use of Technical Analysis in the Foreign Exchange Market. Journal of International Money and Finance, 11, 304-314.

[80] Taylor, M.P., D. Peel and L. Sarno (2001), Non-Linear Mean Reversion in Real Exchange Rates: Towards a Solution to the Purchasing Power Parity Puzzles. International Economic Review, 42, 1015-42.

[81] Taylor, M.P. (2003), Discussion of 'Customer Trades and Extreme Events in Foreign Exchange' and 'Trading Activity, Volatility and Transaction Costs in Spot FX Markets' by M. Fan and R.K. Lyons, in Monetary History, Exchange Rates and Financial Markets: Essays in Honour of Charles Goodhart, P. Mizen (ed.). Edward Elgar Publishing Ltd.

[82] Thomson, P. (2002), Trends in Central Bank Reserve Management. Unpublished JPMorgan Fleming Asset Management Mimeo.

Table 1: Foreign Exchange Market Share Data

Market share in Customer Orders

UBS	11.53%
Citigroup	9.87%
Deutsche Bank	9.79%
JPMorgan Chase	6.79%
Goldman Sachs	5.56%
Credit Suisse First Boston	4.23%
HSBC	3.89%
Morgan Stanley	3.87%
Barclays Capital	3.84%
ABN Amro	3.63%
Merrill Lynch	2.98%
Royal Bank of Scotland	2.85%
Remaining 38 Banks	31.17%

Source: Euromoney (2003)

Notes: Survey covers top 50 banks. Market share data are based on the total volume of foreign exchange business placed annually with each bank. To obtain this figure, respondents to the Euromoney survey estimated the proportion of their total annual foreign exchange dealings placed with their top 10 counterparties. Total business placed with each service provider across all questionnaires received was then divided by total business on all questionnaires ($17.1 trillion) to arrive at a market share figure (Euromoney, 2003).

Table 2: Evans and Lyons (2002a) Model

	No Publication Lag			With Publication Lag	
	Mark	Yen		Mark	Yen
β_1	0.4886	2.6694	β_1	−0.2067	0.7894
	(0.3476)	(0.9189)		(0.5865)	(1.2362)
β_2	2.1498	2.8251	β_2	0.2534	0.4460
	(0.1825)	(0.3539)		(0.3070)	(0.4770)
$Adj.R^2$	0.6353	0.4403	$Adj.R^2$	−0.0157	−0.0105
SSR	0.0004	0.0008	SSR	0.0011	0.0014
AIC	−9.2625	−8.6185	AIC	−8.2251	−8.0415
DW	1.8631	2.2465	DW	2.0400	2.0666
JB^*	0.0151	0.8408	JB^*	0.0000	0.5569
$Breusch^*$	0.5836	0.1531	$Breusch^*$	0.7126	0.6716
$Arch(1)^*$	0.0672	0.9295	$Arch(1)^*$	0.9490	0.1730

Notes: Evans and Lyons (2002a) estimate equation (60) above. Consistent with Evans and Lyons (2002a), estimates of β_2 are multiplied by 10,000. All regressions in this study are estimated with the Newey-West covariance estimator that corrects for the presence of time-varying heteroskedasticity and autocorrelation of unknown form. JB is the Jarque-Bera test for residual normality; Breusch is the Breusch Godfrey test for residual serial correlation. * indicates p-value. Our results differ slightly from Evans and Lyons; we are unable to account for these differences.

Table 3: Granger-causality Tests on Evans and Lyons (2002a) Model

	Order Flow$\Longrightarrow$Exchange Rate	Exchange Rate$\Longrightarrow$Order Flow
Mark	0.1607	0.0084
Yen	0.4093	0.0818

Notes: Table reports p-values for a $\varkappa^2$ test of joint significance of Granger-causality test over 1 to 10 lags. A p-value below 0.05 (0.10) indicates a significant causal relationship at a 5% (10%) level.

Table 4: Perfect Foresight Forecast Errors Using Evans and Lyons (2002a) Interdealer Order Flow

	Horizon	Random Walk	Evans and Lyons	DM
Mark	1 Day	0.4332	0.2848	−1.7887
				(0.0736)
	1 Week	0.9700	0.6248	−1.5032
				(0.1327)
	2 Weeks	1.5162	0.9467	−1.1958
				(0.2317)
Yen	1 Day	0.4018	0.3238	−1.6429
				(0.1004)
	1 Week	0.9513	0.6597	−1.3261
				(0.1847)
	2 Weeks	1.3883	0.8334	−1.4166
				(0.1565)

Notes: The table reports RMSFEs (multiplied by 100) for forecasts at various horizons derived from a random walk model and the Evans and Lyons (2002a) model (equation (60) above). Evans and Lyons forecasts are based upon realised values of the forcing variables, using recursive coefficient estimates starting with the first 39 days of the sample. The final column reports Diebold Mariano (DM) test statistics of forecast accuracy that indicate the significance of differences in forecast accuracy. A negative statistic indicates that the Evans and Lyons model generates more accurate forecasts than the random walk model. Data in parentheses indicate the significance of these differences.

Table 5: Limited Information Forecast Errors Using Evans and Lyons (2002a) Interdealer Order Flow

	Horizon	Random Walk	Evans and Lyons	DM
Mark	1 Day	0.4332	0.5231	2.1681
				(0.0301)
	1 Week	0.9700	1.8073	2.8792
				(0.0039)
	2 Weeks	1.5162	2.7752	3.4880
				(0.0004)
Yen	1 Day	0.4018	0.4191	0.3342
				(0.7381)
	1 Week	0.9513	1.8123	13.2997
				(0.0000)
	2 Weeks	1.3883	3.2625	6.7905
				(0.0000)

Notes: The table reports RMSFEs (multiplied by 100) for forecasts at various horizons derived from a random walk model and the Evans and Lyons (2002a) model (equation (60) above). Evans and Lyons forecasts are based upon realised values of the forcing variables, using recursive coefficient estimates starting with the first 39 days of the sample. The final column reports Diebold Mariano (DM) test statistics of forecast accuracy that indicate the significance of differences in forecast accuracy. A negative statistic indicates that the Evans and Lyons model generates more accurate forecasts than the random walk model. Data in parentheses indicate the significance of these differences.

Table 6: Out-of-Sample Long Horizon Forecasts Using Evans and Lyons (2002a) Interdealer Order Flow

	Horizon (Days)	β_1	β_2
Mark	1	0.3958	0.1992
	2	0.3560	0.1780
	3	0.3178	0.3756
	4	0.1572	0.5000
	5	0.1316	0.3246
	6	0.0674	0.3406
	7	0.0984	0.3418
	8	0.3128	0.3472
	9	0.1980	0.2226
	10	0.2242	0.1516
Yen	1	0.3638	0.2636
	2	0.4830	0.2570
	3	0.3850	0.4844
	4	0.3994	0.3744
	5	0.4200	0.4282
	6	0.3528	0.4324
	7	0.2346	0.4826
	8	0.2166	0.3564
	9	0.3320	0.2962
	10	0.4248	0.2376

Notes: The table reports p-values associated with coefficient estimates from the Evans and Lyons (2002a) model (equation (60) above). P-values are derived from a non-parametric bootstrap that runs 5000 monte carlo simulations of a naive random walk model over each forecast horizon from i = 1 to 10 and indicate the significance of observed t-statistics from the Evans and Lyons (2002a) model. The F-stat is a χ^2 test of joint significance of the estimated coefficients. A value equal to or less than 0.05 indicates a significant t-statistic.

Table 7: JPM Aggregate Customer Order Flow Data Correlation Analysis

		Correlation of Daily Exchange Rate Returns with:	
		Net Inflows into Domestic Currency	Net Inflows into US Dollar
Euro	1999	−0.1794	0.1532
	2000	−0.1364	0.0195
	2001	−0.2048	0.0869
	2002	−0.1171	0.1673
	2003	−0.0355	0.2093
Yen	1999	−0.1290	0.0756
	2000	−0.1747	0.0946
	2001	−0.1236	0.1744
	2002	−0.0765	0.1634
	2003	−0.1899	0.1885
Sterling	1999	0.0495	0.0441
	2000	−0.0907	0.1494
	2001	−0.1706	0.1043
	2002	−0.0015	0.1619
	2003	−0.0886	0.1748
Swiss franc	1999	−0.1521	0.1565
	2000	−0.1025	0.0428
	2001	−0.1193	0.0962
	2002	−0.0338	0.1992
	2003	−0.2362	0.2290

Table 8: Evans and Lyons (2002a) Model Using JPM Aggregate Customer Data

	No Publication Lag			With Publication Lag	
	Euro	Yen		Euro	Yen
β_1	0.2736	0.0896	β_1	-0.2792	-0.2308
	(0.2867)	(0.2514)		(0.3101)	(0.2530)
β_2	-0.0992	-0.0671	β_2	0.0025	0.0274
	(0.0263)	(0.0367)		(0.0265)	(0.0335)
β_3	0.0491	0.0765	β_3	0.0062	0.0045
	(0.0285)	(0.0346)		(0.0292)	(0.0346)
$Adj.R^2$	0.0219	0.0169	$Adj.R^2$	-0.0008	-0.0003
SSR	0.0448	0.0481	SSR	0.0459	0.0489
AIC	-7.3083	-7.2372	AIC	-7.2844	-7.2206
DW	1.9821	2.0220	DW	1.9511	2.0022
JB^*	0.0000	0.0000	JB^*	0.0000	0.0000
$Breusch^*$	0.8124	0.6614	$Breusch^*$	0.4436	0.9145
$Arch(1)^*$	0.8513	0.0641	$Arch(1)^*$	0.7933	0.2422

Notes: We re-estimate the Evans and Lyons (2002a) model (equation (60) above), where ΔS_t, $\Delta(i_t - i_t^*)$ and η_t are defined as above. X_t is aggregate customer order flow transacted with JPM as measured by the FXI between 4 p.m. GMT on day $t-1$ and the same time on day t. The estimate of β_2 is multiplied by 10,000. * signifies p-value.

Table 8 (cont.): Evans and Lyons (2002a) Model Using JPM Aggregate Customer Data

	No Publication Lag			With Publication Lag	
	Sterling	Swiss franc		Sterling	Swiss franc
β_1	−0.1605	0.3541	β_1	−0.3871	−0.0921
	(0.1957)	(0.3105)		(0.2004)	(0.3246)
β_2	−0.0329	−0.8697	β_2	0.0137	0.0240
	(0.0211)	(0.3080)		(0.0218)	(0.3003)
β_3	0.0867	0.9176	β_3	0.0388	0.2487
	(0.0222)	(0.3027)		(0.0216)	(0.2963)
$Adj.R^2$	0.0208	0.0233	$Adj.R^2$	0.0050	−0.0010
SSR	0.0244	0.0470	SSR	0.0248	0.0482
AIC	−7.9159	−7.2610	AIC	−7.8993	−7.2355
DW	1.9964	2.0226	DW	1.9804	2.0030
JB^*	0.0000	0.0000	JB^*	0.0000	0.0000
$Breusch^*$	0.9716	0.6419	$Breusch^*$	0.7532	0.8529
$Arch(1)^*$	0.8205	0.5415	$Arch(1)^*$	0.7455	0.4307

Notes: We re-estimate the Evans and Lyons (2002a) model (equation (60) above), where ΔS_t, $\Delta(i_t - i_t^*)$ and η_t are defined as above. X_t is aggregate customer order flow transacted with JPM as measured by the FXI between 4 p.m. GMT on day $t-1$ and the same time on day t. The estimate of β_2 is multiplied by 10,000. * signifies p-value.

Table 9: Granger-causality Tests on Evans and Lyons (2002a) Model Using JPM Aggregate Customer Data

	Null Hypothesis	P-Value
Euro	$DCCYORDER \Longrightarrow DLCCY$	0.6863
	$DLCCY \Longrightarrow DCCYORDER$	0.0000
	$DUSORDER \Longrightarrow DLCCY$	0.4640
	$DLCCY \Longrightarrow DUSORDER$	0.0000
Yen	$DCCYORDER \Longrightarrow DLCCY$	0.0108
	$DLCCY \Longrightarrow DCCYORDER$	0.0000
	$DUSORDER \Longrightarrow DLCCY$	0.2525
	$DLCCY \Longrightarrow DUSORDER$	0.0000
Sterling	$DCCYORDER \Longrightarrow DLCCY$	0.2842
	$DLCCY \Longrightarrow DCCYORDER$	0.0024
	$DUSORDER \Longrightarrow DLCCY$	0.1302
	$DLCCY \Longrightarrow DUSORDER$	0.0000
Swiss franc	$DCCYORDER \Longrightarrow DLCCY$	0.5959
	$DLCCY \Longrightarrow DCCYORDER$	0.0000
	$DUSORDER \Longrightarrow DLCCY$	0.5151
	$DLCCY \Longrightarrow DUSORDER$	0.0000

Notes: Table reports p-values for a $\varkappa^2$ test of joint significance of Granger-causality test over 1 to 10 lags. A p-value below 0.05 (0.10) indicates a significant causal relationship at a 5% (10%) level.

Table 10: Perfect Foresight Forecast Errors Using JPM Aggregate Customer Data

	Horizon	Random Walk	Evans and Lyons	DM
Euro	1 Day	0.6470	0.6418	−1.7453
				(0.0809)
	1 Week	1.4726	1.4282	−2.3614
				(0.0182)
	2 Weeks	2.1156	2.0258	−2.2163
				(0.0266)
Yen	1 Day	0.6146	0.6152	0.1500
				(0.8807)
	1 Week	1.3170	1.2954	−0.6817
				(0.4953)
	2 Weeks	1.8842	1.8103	−1.2379
				(0.2157)
Sterling	1 Day	0.4763	0.4724	−1.6527
				(0.0983)
	1 Week	1.0725	1.0228	−2.8603
				(0.0042)
	2 Weeks	1.5303	1.4034	−3.3240
				(0.0008)
Swiss franc	1 Day	0.6560	0.6527	−0.8871
				(0.3749)
	1 Week	1.4626	1.4148	−1.8867
				(0.0592)
	2 Weeks	2.0928	1.9786	−2.0862
				(0.0369)

Notes: The table reports RMSFEs (multiplied by 100) for forecasts derived from a random walk and the Evans and Lyons (2002a) model using JPM FXI order flow. Evans and Lyons forecasts are based upon realised values of the forcing variables using recursive coefficient estimates starting with the initial 200 days of the sample. A negative DM test statistic indicates that Evans and Lyons forecasts are more accurate than a random walk. Data in parentheses indicate the significance of these differences.

Table 11: Limited Information Forecast Errors Using JPM Aggregate Customer Data

	Horizon	Random Walk	Evans and Lyons	DM
Euro	1 Day	0.6470	0.6547	2.7860
				(0.0053)
	1 Week	1.4726	1.5597	3.3982
				(0.0006)
	2 Weeks	2.1156	2.3406	4.1811
				(0.0000)
Yen	1 Day	0.6146	0.6259	3.1035
				(0.0019)
	1 Week	1.3170	1.4678	3.9389
				(0.0000)
	2 Weeks	1.8804	2.2768	4.9327
				(0.0000)
Sterling	1 Day	0.4763	0.4769	0.2570
				(0.7971)
	1 Week	1.0725	1.0888	0.8232
				(0.4103)
	2 Weeks	1.5303	1.6232	2.1234
				(0.0337)
Swiss franc	1 Day	0.6560	0.6631	1.8351
				(0.0664)
	1 Week	1.4626	1.5678	3.4276
				(0.0006)
	2 Weeks	2.0928	2.3887	4.8372
				(0.0000)

Notes: The table reports RMSFEs (multiplied by 100) for forecasts derived from a random walk and the Evans and Lyons (2002a) model using JPM FXI order flow. Evans and Lyons forecasts are based upon realised values of the forcing variables using recursive coefficient estimates starting with the initial 200 days of the sample. A negative DM test statistic indicates that Evans and Lyons forecasts are more accurate than a random walk. Data in parentheses indicate the significance of these differences.

Table 12: Long Horizon Forecasts Using JPM Aggregate Customer Data

	Horizon (Days)	β_1	β_2	β_3	$F-Stat$
Euro	1	0.1896	0.4470	0.4174	0.8420
	2	0.1728	0.3432	0.4864	0.7890
	3	0.1178	0.2502	0.4732	0.5844
	4	0.0274	0.1850	0.4956	0.2016
	5	0.1978	0.1948	0.4744	0.6686
	6	0.1700	0.2318	0.4670	0.6344
	7	0.4372	0.2906	0.4004	0.8856
	8	0.3466	0.3410	0.3852	0.8830
	9	0.4706	0.3566	0.3964	0.9500
	10	0.2944	0.3600	0.3690	0.8554
Yen	1	0.1886	0.2266	0.4486	0.6732
	2	0.2220	0.1938	0.4568	0.6562
	3	0.3428	0.0744	0.3654	0.3664
	4	0.3532	0.0746	0.3098	0.3982
	5	0.2082	0.0618	0.3214	0.2808
	6	0.4906	0.0440	0.2594	0.3058
	7	0.4526	0.0362	0.2518	0.2748
	8	0.4142	0.0440	0.3002	0.2878
	9	0.3336	0.0398	0.3432	0.2392
	10	0.1762	0.0346	0.3710	0.1658

Notes: The table reports p-values associated with coefficient estimates from the Evans and Lyons (2002a) model (equation (60) above) using JPM customer order data. P-values are derived from a non-parametric bootstrap that runs 5000 monte carlo simulations of a naive random walk model over each forecast horizon from i = 1 to 10 and indicate the significance of observed t-statistics from the Evans and Lyons (2002a) model. A value equal to or less than 0.05 indicates a significant t-statistic. The F-stat is a χ^2 test of joint significance of the estimated coefficients. The number of simulations at each forecast horizon is adjusted where necessary to account for presence of collinearity. Details available on request.

Table 12 (cont.): Long Horizon Forecasts Using JPM Aggregate Customer Data

	Horizon (Days)	β_1	β_2	β_3	$F - Stat$
Sterling	1	0.0224	0.2630	0.0326	0.0712
	2	0.1214	0.2920	0.0354	0.2150
	3	0.0642	0.2668	0.0212	0.0992
	4	0.0252	0.1872	0.0116	0.0360
	5	0.0412	0.1756	0.0106	0.0662
	6	0.0844	0.1374	0.0042	0.0590
	7	0.1356	0.0798	0.0086	0.0668
	8	0.1054	0.0634	0.0158	0.1034
	9	0.1110	0.0780	0.0348	0.1794
	10	0.0318	0.0880	0.0640	0.1266
Swiss franc	1	0.3858	0.4660	0.2202	0.8322
	2	0.0686	0.1804	0.1600	0.3342
	3	0.0748	0.1306	0.2066	0.3084
	4	0.1526	0.1504	0.2358	0.4374
	5	0.3088	0.1026	0.2346	0.5156
	6	0.2854	0.0776	0.2814	0.4700
	7	0.4800	0.0686	0.3692	0.4992
	8	0.3504	0.0718	0.3936	0.4754
	9	0.4546	0.0560	0.3778	0.4484
	10	0.4072	0.0660	0.4222	0.4616

Notes: The table reports p-values associated with coefficient estimates from the Evans and Lyons (2002a) model (equation (60) above) using JPM customer order data. P-values are derived from a non-parametric bootstrap that runs 5000 monte carlo simulations of a naive random walk model over each forecast horizon from i = 1 to 10 and indicate the significance of observed t-statistics from the Evans and Lyons (2002a) model. A value equal to or less than 0.05 indicates a significant t-statistic. The F-stat is a χ^2 test of joint significance of the estimated coefficients. The number of simulations at each forecast horizon is adjusted where necessary to account for presence of collinearity. Details available on request.

Table 13: RBS Aggregate Customer Order Flow Data Correlation Analysis

		Correlation with Daily Exchange Rate Returns:
Euro	1st Half	-0.0160
	2nd Half	0.1056
Yen	1st Half	0.0657
	2nd Half	0.2058
Sterling	1st Half	-0.0298
	2nd Half	−0.0427
Swiss franc	1st Half	0.0787
	2nd Half	0.1555

Notes: Data for July 2002 was not collected. Consequently, correlation coefficients are calculated over two halves of sample. A negative correlation indicates that net orders of currency i are associated with an appreciation of this currency.

Table 14: Evans and Lyons (2002a) Model Applied to RBS Aggregate Customer Data

	No Publication Lag			With Publication Lag	
	Euro	Yen		Euro	Yen
β_1	−0.4796	−0.9152	β_1	−0.5022	0.6539
	(0.7133)	(0.8989)		(0.7092)	(0.9291)
β_2	−64.3671	488.9440	β_2	−155.7218	99.8080
	(142.8075)	(155.0039)		(142.9603)	(157.6374)
$Dum1$	−0.0038	0.0023	$Dum1$	−0.0041	0.0001
	(0.0055)	(0.0058)		(0.0055)	(0.0058)
$Dum2$	0.0051	−0.0028	$Dum2$	0.0054	−0.0024
	(0.0055)	(0.0058)		(0.0055)	(0.0059)
$Adj.R^2$	−0.0138	0.0217	$Adj.R^2$	−0.0111	−0.0053
SSR	0.0116	0.0128	SSR	0.0116	0.0131
AIC	−7.5531	−7.4539	AIC	−7.5533	−7.4268
DW	2.0025	2.0985	DW	2.0082	2.0524
JB^*	0.0002	0.0000	JB^*	0.0004	0.0000
$Breusch^*$	0.9843	0.3195	$Breusch^*$	0.9258	0.6051
$Arch(1)^*$	0.7691	0.5140	$Arch(1)^*$	0.8339	0.4737

Notes: We re-estimate equation (60) above, where ΔS_t, $\Delta(i_t - i_t^*)$ and η_t are defined as above . X_t is the flow of net customer purchases of currency i transacted through RBS from 4 p.m. GMT on day $t-1$ to the same time on day t divided by the gross volume of flows. Data calculated as an index, and constrained to lie within values +/-1.0. Dum1 and Dum2 are dummy variables used as a splice for omitted data in July 2002. Estimates of β_2 are multiplied by 10,000. * signifies p-value.

Table 14 (Cont.): Evans and Lyons (2002a) Model Applied to RBS Aggregate Customer Data

	No Publication Lag			With Publication Lag	
	Sterling	Swiss franc		Sterling	Swiss franc
β_1	−0.3944	0.1752	β_1	0.1743	−0.3008
	(0.4063)	(0.7166)		(0.4067)	(0.7229)
β_2	−215.0122	157.1606	β_2	−129.8212	−63.3646
	(121.6097)	(86.9848)		(121.4232)	(87.0259)
*Dum*1	−0.0200	-0.0073	*Dum*1	−0.0200	−0.0083
	(0.0044)	(0.0066)		(0.0043)	(0.0060)
*Dum*2	−0.0083	0.0001	*Dum*2	−0.0086	−0.0022
	(0.0043)	(0.0060)		(0.0044)	(0.0066)
$Adj.R^2$	0.0626	−0.0020	$Adj.R^2$	0.0547	−0.0089
SSR	0.0073	0.0140	*SSR*	0.0073	0.0141
AIC	−8.0097	−7.3644	*AIC*	−8.0114	−7.3549
DW	1.9510	2.0039	*DW*	1.9352	2.0026
JB^*	0.0000	0.0000	JB^*	0.0000	0.0000
*Breusch**	0.7297	0.9682	*Breusch**	0.5735	0.9604
$Arch(1)^*$	0.4884	0.9776	$Arch(1)^*$	0.4671	0.8855

Notes: We re-estimate equation (60) above, where ΔS_t, $\Delta(i_t - i_t^*)$ and η_t are defined as above. X_t is the flow of net customer purchases of currency i transacted through RBS from 4 p.m. GMT on day $t-1$ to the same time on day t divided by the gross volume of flows. Data calculated as an index, and constrained to lie within values +/-1.0. Dum1 and Dum2 are dummy variables used as a splice for omitted data in July 2002. Estimates of β_2 are multiplied by 10,000. * signifies p-value.

Table 15: Granger-causality Tests on Evans and Lyons (2002a) Model Using RBS Aggregate Customer Data

	Null Hypothesis	1st Half P-Value	2nd Half P-Value
Euro	$DORDER \Longrightarrow DLCCY$	0.0841	0.5468
	$DLCCY \Longrightarrow DORDER$	0.0728	0.2417
Yen	$DORDER \Longrightarrow DLCCY$	0.0698	0.2603
	$DLCCY \Longrightarrow DORDER$	0.0142	0.0000
Sterling	$DORDER \Longrightarrow DLCCY$	0.2807	0.2135
	$DLCCY \Longrightarrow DORDER$	0.6146	0.3024
Swiss franc	$DORDER \Longrightarrow DLCCY$	0.6232	0.3733
	$DLCCY \Longrightarrow DORDER$	0.3238	0.0624

Notes: Table reports p-values for a $\varkappa^2$ test of joint significance of Granger-causality test over 1 to 10 lags. A p-value below 0.05 (0.10) indicates a significant causal relationship at a 5% (10%) level.

Table 16: Perfect Foresight Forecast Errors Using RBS Aggregate Customer Data

	Horizon	Random Walk	Evans and Lyons	DM
Euro	1 Day	0.5596	0.5606	0.6456
				(0.5184)
	1 Week	1.3722	1.3694	−0.3911
				(0.6957)
	2 Weeks	1.9256	1.9206	−0.3324
				(0.7395)
Yen	1 Day	0.5986	0.5898	−1.5394
				(0.1237)
	1 Week	1.3230	1.2592	−2.4243
				(0.0153)
	2 Weeks	1.7941	1.6724	−2.8647
				(0.0041)
Sterling	1 Day	0.4854	0.4861	0.1812
				(0.8561)
	1 Week	1.0793	1.0921	1.1403
				(0.2541)
	2 Weeks	1.4949	1.5183	1.1601
				(0.2459)
Swiss franc	1 Day	0.6223	0.6239	0.3895
				(0.6968)
	1 Week	1.4639	1.4483	−0.8500
				(0.3952)
	2 Weeks	2.0820	2.0572	−0.6411
				(0.5214)

Notes: The table reports RMSFEs (multiplied by 100) for forecasts derived from a random walk and the Evans and Lyons (2002a) model using RBS aggregate order flow. Evans and Lyons forecasts are based upon realised values of the forcing variables using recursive coefficient estimates starting with the initial 200 days of the sample. A negative DM test statistic indicates that Evans and Lyons forecasts are more accurate than a random walk. Data in parentheses indicate the significance of these differences.

Table 17: Limited Information Forecast Errors Using RBS Aggregate Customer Data

	Horizon	Random Walk	Evans and Lyons	DM
Euro	1 Day	0.5596	0.5581	−0.8476
				(0.3966)
	1 Week	1.3722	1.3857	1.1816
				(0.2373)
	2 Weeks	1.9256	1.9635	1.3452
				(0.1785)
Yen	1 Day	0.5986	0.5986	0.8382
				(0.4019)
	1 Week	1.3230	1.4056	2.3391
				(0.0193)
	2 Weeks	1.7941	2.0368	4.3961
				(0.0000)
Sterling	1 Day	0.4854	0.4875	0.6456
				(0.5184)
	1 Week	1.0793	1.1164	2.2430
				(0.0248)
	2 Weeks	1.4949	1.6064	3.6624
				(0.0002)
Swiss franc	1 Day	0.6223	0.6231	0.2805
				(0.7790)
	1 Week	1.4639	1.4712	0.1778
				(0.8588)
	2 Weeks	2.0820	2.1330	1.4781
				(0.1393)

Notes: The table reports RMSFEs (multiplied by 100) for forecasts derived from a random walk and the Evans and Lyons (2002a) model using RBS aggregate order flow. Evans and Lyons forecasts are based upon realised values of the forcing variables using recursive coefficient estimates starting with the initial 200 days of the sample. A negative DM test statistic indicates that Evans and Lyons forecasts are more accurate than a random walk. Data in parentheses indicate the significance of these differences.

Table 18: Long Horizon Forecast Errors Using RBS Aggregate Customer Data

	Horizon (Days)	β_1	β_2	$F - Stat$
Euro	1	0.2256	0.1686	0.4438
	2	0.3966	0.0524	0.2430
	3	0.4796	0.4700	0.9546
	4	0.4418	0.2518	0.6746
	5	0.3476	0.1734	0.4776
	6	0.3870	0.1000	0.3159
	7	0.2456	0.1406	0.3608
	8	0.2708	0.1014	0.3004
	9	0.1314	0.0878	0.1624
	10	0.0938	0.0886	0.1397
Yen	1	0.3038	0.2418	0.7108
	2	0.3176	0.2666	0.7606
	3	0.2912	0.0828	0.3972
	4	0.4484	0.0716	0.4158
	5	0.4560	0.1306	0.5567
	6	0.4682	0.1382	0.6117
	7	0.3860	0.1230	0.5703
	8	0.2708	0.1474	0.5286
	9	0.072	0.2850	0.3735
	10	0.2814	0.0522	0.2402

Notes: The table reports p-values associated with coefficient estimates derived from the Evans and Lyons (2002a) model (equation (60) above) using RBS aggregate customer order data. P-values calculated as above, but excluding the initial 194 sample observations. The F-stat is a χ^2 test of joint significance of the estimated coefficients. The number of simulations at each forecast horizon is adjusted where necessary to account for presence of collinearity. Details available on request.

Table 18 (Cont.): Long Horizon Forecast Errors Using RBS Aggregate Customer Data

	Horizon (Days)	β_1	β_2	$F-Stat$
Sterling	1	0.3180	0.1160	0.4994
	2	0.4458	0.3354	0.9080
	3	0.2562	0.1678	0.5756
	4	0.2236	0.0962	0.4042
	5	0.2094	0.2344	0.6104
	6	0.1820	0.1832	0.5299
	7	0.2052	0.3484	0.6553
	8	0.1500	0.4826	0.5372
	9	0.2040	0.4196	0.6706
	10	0.1068	0.3104	0.3797
Swiss franc	1	0.3458	0.1934	0.7198
	2	0.3058	0.5822	0.9116
	3	0.1838	0.3296	0.5388
	4	0.3048	0.3540	0.7088
	5	0.2284	0.1430	0.3274
	6	0.2148	0.1996	0.4342
	7	0.2030	0.2362	0.4970
	8	0.1608	0.2282	0.3874
	9	0.2480	0.3526	0.6126
	10	0.3226	0.4020	0.7123

Notes: The table reports p-values associated with coefficient estimates derived from the Evans and Lyons (2002a) model (equation (60) above) using RBS aggregate customer order data. P-values calculated as above, but excluding the initial 194 sample observations. The F-stat is a χ^2 test of joint significance of the estimated coefficients. The number of simulations at each forecast horizon is adjusted where necessary to account for presence of collinearity. Details available on request.

Table 19: RBS Disaggregated Customer Order Flow Data Correlation Analysis

		Correlation of Exchange Rate Returns with:			
		Corporate	Real	Leveraged	Other
Euro	1st Half	0.0247	0.0654	−0.0420	−0.0882
	2nd Half	0.0276	0.0398	0.0763	−0.0250
Yen	1st Half	0.0516	−0.0227	−0.0178	0.0722
	2nd Half	0.0781	0.2446	0.1831	−0.0079
Sterling	1st Half	0.0510	−0.0790	−0.0281	−0.1163
	2nd Half	0.0471	−0.1608	0.0086	−0.0678
Swiss franc	1st Half	0.0531	−0.0523	0.1278	0.0698
	2nd Half	0.0206	0.0897	0.0786	0.1231

Notes: Data for July 2002 was not collected. Consequently, correlation coefficients calculated over two halves of sample. A negative correlation indicates that net orders of currency i are associated with an appreciation of this currency.

Table 20: Evans and Lyons (2002a) Model Applied to RBS Disaggregated Customer Data

	No Publication Lag			With Publication Lag	
	Euro	Yen		Euro	Yen
β_1	−0.5039	−0.8023	β_1	−0.5323	0.6003
	(0.7609)	(0.7666)		(0.6343)	(1.1922)
β_2	18.9184	−97.6561	β_2	5.3248	24.2875
	(79.1380)	(75.9510)		(80.5618)	(81.9591)
β_3	−6.3183	188.4896	β_3	−49.5647	29.3428
	(63.3373)	(60.7800)		(61.4913)	(61.1588)
β_4	−54.0679	102.6600	β_4	−28.3571	91.9986
	(40.6930)	(37.8056)		(41.4562)	(44.0591)
β_5	−70.7078	218.9518	β_5	−107.8135	−47.2633
	(119.1708)	(105.1958)		(124.0523)	(115.3405)
$Dum1$	−0.0039	0.0016	$Dum1$	−0.0038	−0.0005
	(0.0004)	(0.0006)		(0.0004)	(0.0004)
$Dum2$	0.0049	−0.0043	$Dum2$	0.0051	−0.0020
	(0.0006)	(0.0005)		(0.0004)	(0.0006)
$Adj.R^2$	−0.0168	0.0440	$Adj.R^2$	−0.0177	0.0008
SSR	0.0115	0.0124	SSR	0.0115	0.0129
AIC	−8.0522	−7.4697	AIC	−7.5392	−7.5392
DW	2.0215	2.1600	DW	2.0021	2.0560
JB^*	0.0004	0.0000	JB^*	0.0008	0.0000
$Breusch^*$	0.0000	0.1002	$Breusch^*$	0.9745	0.0000
$Arch(1)^*$	0.2145	0.6180	$Arch(1)^*$	0.7988	0.4677

Notes: We estimate a version of equation (60) above. X_t is the flow of net customer purchases of currency i transacted through RBS disaggregated into the following sub-components: net corporate (OC), real money (OR), leveraged money (OL) and other customer orders (OO). Dum1 and Dum2 are dummy variables used as a splice for omitted data in July 2002. Our regression is

$$\Delta S_t = \beta_1 \Delta(i_t - i_t^*) + \beta_2 OC_t + \beta_3 OR_t + \beta_4 OL_t + \beta_5 OO_t + Dum1 + Dum2 + \eta_t \quad (63)$$

Estimates of $\beta_2 ... \beta_5$ are multiplied by 10,000. * signifies p-value.

Table 20 (cont.): Evans and Lyons (2002a) Model Applied to RBS Disaggregated Customer Data

	No Publication Lag			With Publication Lag	
	Sterling	Swiss franc		Sterling	Swiss franc
β_1	−0.4998	−0.0552	β_1	0.1478	−0.2765
	(0.5123)	(0.7704)		(0.3890)	(0.6653)
β_2	191.5454	−99.2095	β_2	−9.9745	−115.7387
	(77.9776)	(50.8914)		(91.1450)	(55.1375)
β_3	−125.2563	20.3724	β_3	−40.0325	−51.5165
	(44.3892)	(43.0460)		(49.7471)	(42.9183)
β_4	−93.5831	97.4216	β_4	−9.0412	21.8917
	(28.4791)	(41.3283)		(29.9371)	(43.5395)
β_5	−107.9147	208.9912	β_5	−50.9129	−16.4408
	(82.4005)	(64.4209)		(92.4954)	(75.8017)
*Dum*1	−0.0217	−0.0065	*Dum*1	−0.0200	−0.0093
	(0.0007)	(0.0028)		(0.0003)	(0.0005)
*Dum*2	−0.0074	−0.0004	*Dum*2	−0.0088	−0.0013
	(0.0004)	(0.0006)		(0.0006)	(0.0024)
$Adj.R^2$	0.1053	0.0328	$Adj.R^2$	0.0471	−0.0036
SSR	0.0069	0.0134	*SSR*	0.0073	0.0139
AIC	−8.0522	−7.3921	*AIC*	−7.9957	−7.3526
DW	2.0397	2.1198	*DW*	1.9434	2.0128
JB^*	0.0000	0.0000	JB^*	0.0000	0.0000
*Breusch**	0.6029	0.2309	*Breusch**	0.6185	0.8795
$Arch(1)^*$	0.4799	0.9564	$Arch(1)^*$	0.4751	0.9974

Notes: We estimate a version of equation (4) above. X_t is the flow of net customer purchases of currency i transacted through RBS disaggregated into the following sub-components: net corporate (OC), real money (OR), leveraged money (OL) and other customer orders (OO). Dum1 and Dum2 are dummy variables used as a splice for omitted data in July 2002. Estimates of $\beta_2 ... \beta_5$ are multiplied by 10,000. * signifies p-value.

Table 21: Granger-causality Tests on Evans and Lyons (2002a) Model Using RBS Disaggregated Customer Data

	Null Hypothesis	1st Half P-Value	2nd Half P-Value
Euro	$OC \Longrightarrow DLCCY$	0.3617	0.8347
	$DLCCY \Longrightarrow OC$	0.6712	0.0910
	$OR \Longrightarrow DLCCY$	0.9415	0.5001
	$DLCCY \Longrightarrow OR$	0.0674	0.4770
	$OL \Longrightarrow DLCCY$	0.2069	0.0995
	$DLCCY \Longrightarrow OL$	0.4030	0.5280
	$OO \Longrightarrow DLCCY$	0.0363	0.4949
	$DLCCY \Longrightarrow OO$	0.1473	0.0001
Yen	$OC \Longrightarrow DLCCY$	0.2362	0.7086
	$DLCCY \Longrightarrow OC$	0.0067	0.0129
	$OR \Longrightarrow DLCCY$	0.1531	0.0567
	$DLCCY \Longrightarrow OR$	0.5821	0.0220
	$OL \Longrightarrow DLCCY$	0.0667	0.7300
	$DLCCY \Longrightarrow OL$	0.0002	0.0000
	$OO \Longrightarrow DLCCY$	0.1877	0.6873
	$DLCCY \Longrightarrow OO$	0.3484	0.4125

Notes: Table reports p-values for a $\varkappa^2$ test of joint significance of Granger-causality test over 1 to 10 lags. A p-value below 0.05 (0.10) indicates a significant causal relationship at a 5% (10%) level.

Table 21 (cont.): Granger-causality Tests on Evans and Lyons (2002a) Model Using RBS Disaggregated Customer Data

	Null Hypothesis	1st Half P-Value	2nd Half P-Value
Sterling	$OC \Longrightarrow DLCCY$	0.1543	0.3062
	$DLCCY \Longrightarrow OC$	0.0634	0.0003
	$OR \Longrightarrow DLCCY$	0.1852	0.4805
	$DLCCY \Longrightarrow OR$	0.7994	0.0398
	$OL \Longrightarrow DLCCY$	0.6921	0.3953
	$DLCCY \Longrightarrow OL$	0.1538	0.0080
	$OO \Longrightarrow DLCCY$	0.3395	0.0010
	$DLCCY \Longrightarrow OO$	0.0018	0.1108
Swiss franc	$OC \Longrightarrow DLCCY$	0.1366	0.0001
	$DLCCY \Longrightarrow OC$	0.0000	0.1797
	$OR \Longrightarrow DLCCY$	0.4014	0.6113
	$DLCCY \Longrightarrow OR$	0.7720	0.7577
	$OL \Longrightarrow DLCCY$	0.6714	0.6639
	$DLCCY \Longrightarrow OL$	0.0012	0.0110
	$OO \Longrightarrow DLCCY$	0.1865	0.1035
	$DLCCY \Longrightarrow OO$	0.0184	0.0046

Notes: Table reports p-values for a $\varkappa^2$ test of joint significance of Granger-causality test over 1 to 10 lags. A p-value below 0.05 (0.10) indicates a significant causal relationship at a 5% (10%) level.

Table 22: Perfect Foresight Forecast Errors Using RBS Disaggregated Customer Data

	Horizon	Random Walk	Evans and Lyons	DM
Euro	1 Day	0.5596	0.5666	1.6797
				(0.0929)
	1 Week	1.3722	1.3616	−0.5360
				(0.5919)
	2 Weeks	1.9256	1.8845	−0.9897
				(0.3222)
Yen	1 Day	0.5986	0.5828	−1.6139
				(0.1065)
	1 Week	1.3230	1.2040	−3.4285
				(0.0006)
	2 Weeks	1.7941	1.5414	−5.1429
				(0.0000)
Sterling	1 Day	0.4854	0.4718	−2.2585
				(0.0239)
	1 Week	1.0793	1.0046	−3.3842
				(0.0007)
	2 Weeks	1.4949	1.3364	−3.7016
				(0.0002)
Swiss franc	1 Day	0.6223	0.6097	−1.5314
				(0.1256)
	1 Week	1.4639	1.3194	−3.9277
				(0.0000)
	2 Weeks	2.0820	1.8205	−3.8423
				(0.0001)

Notes: The table reports RMSFEs (multiplied by 100) for forecasts derived from a random walk and the Evans and Lyons (2002a) model using RBS disaggregated customer order data. Evans and Lyons forecasts are based upon realised values of the forcing variables using recursive coefficient estimates starting with the initial 200 days of the sample. A negative DM test statistic indicates that Evans and Lyons forecasts are more accurate than a random walk. Data in parentheses indicate the significance of these differences.

Table 23: Limited Information Forecast Errors Using RBS Disaggregated Customer Data

	Horizon	Random Walk	Evans and Lyons	DM
Euro	1 Day	0.5596	0.5621	0.6345
				(0.5257)
	1 Week	1.3722	1.3835	0.4974
				(0.6188)
	2 Weeks	1.9256	1.9805	1.3219
				(0.1861)
Yen	1 Day	0.5986	0.6124	1.2219
				(0.2217)
	1 Week	1.3230	1.5237	3.4360
				(0.0005)
	2 Weeks	1.7941	2.3605	5.5638
				(0.0000)
Sterling	1 Day	0.4854	0.4933	1.3782
				(0.1681)
	1 Week	1.0793	1.1573	2.5332
				(0.0112)
	2 Weeks	1.4949	1.7450	3.9937
				(0.0000)
Swiss franc	1 Day	0.6223	0.6101	−1.5114
				(0.1306)
	1 Week	1.4639	1.5154	1.1068
				(0.2683)
	2 Weeks	2.0820	2.2854	1.7259
				(0.0843)

Notes: The table reports RMSFEs (multiplied by 100) for forecasts derived from a random walk and the Evans and Lyons (2002a) model using RBS disaggregated customer order data. Evans and Lyons forecasts are based upon realised values of the forcing variables using recursive coefficient estimates starting with the initial 200 days of the sample. A negative DM test statistic indicates that Evans and Lyons forecasts are more accurate than a random walk. Data in parentheses indicate the significance of these differences.

Table 24: Long Horizon Forecast Errors Using RBS Disaggregated Customer Data

	Horizon (Days)	β_1	β_2	β_3	β_4	β_5	$F - stat$
Euro	1	0.1992	0.4694	0.1076	0.2970	0.2198	0.7960
	2	0.3908	0.2158	0.1468	0.5414	0.3394	0.9056
	3	0.4814	0.4360	0.5336	0.3178	0.4388	0.9638
	4	0.4470	0.2946	0.2366	0.2166	0.1894	0.8245
	5	0.3500	0.3204	0.7162	0.1944	0.1642	0.7597
	6	0.4208	0.3678	0.1372	0.2578	0.3026	0.9257
	7	0.2714	0.3852	0.1352	0.3662	0.3992	0.9909
	8	0.2868	0.5020	0.1460	0.2266	0.1950	0.8874
	9	0.1272	0.2328	0.6020	0.2992	0.1500	0.9108
	10	0.0908	0.3088	0.5158	0.1896	0.0734	0.6741
Yen	1	0.3172	0.3910	0.3232	0.0202	0.3480	0.2578
	2	0.3186	0.1488	0.4642	0.2714	0.4540	0.8516
	3	0.2856	0.0408	0.3564	0.2118	0.3002	0.4708
	4	0.4338	0.0236	0.3050	0.3102	0.3110	0.4009
	5	0.4464	0.0466	0.2142	0.2682	0.4766	0.5405
	6	0.4558	0.0880	0.2982	0.3064	0.3270	0.7886
	7	0.3914	0.0918	0.3600	0.3696	0.3170	0.8448
	8	0.2774	0.1130	0.5192	0.2724	0.2934	0.9026
	9	0.2998	0.1020	0.3072	0.4192	0.2154	0.9105
	10	0.3094	0.0644	0.3424	0.4442	0.1046	0.7028

Notes: The table reports p-values associated with coefficient estimates derived from the Evans and Lyons (2002a) model (equation (60) above) using RBS aggregate customer order data. P-values calculated as above, but excluding the initial 194 sample observations. The F-stat is a χ^2 test of joint significance of the estimated coefficients. The number of simulations at each forecast horizon is adjusted where necessary to account for presence of collinearity. Details available on request.

Table 24 (Cont.): Long Horizon Forecast Errors Using RBS Disaggregated Customer Data

	Horizon (Days)	β_1	β_2	β_3	β_4	β_5	$F - stat$
Sterling	1	0.3418	0.4252	0.1850	0.4850	0.2292	0.8040
	2	0.3362	0.3828	0.3210	0.3080	0.2820	0.9262
	3	0.2610	0.2668	0.4044	0.4530	0.1340	0.8253
	4	0.2394	0.3846	0.2962	0.5246	0.0746	0.6701
	5	0.2124	0.5096	0.3282	0.4902	0.2200	0.9450
	6	0.1764	0.4018	0.1790	0.5822	0.2276	0.9078
	7	0.1972	0.5524	0.3244	0.2072	0.1886	0.9294
	8	0.1294	0.4080	0.4968	0.0812	0.1638	0.8148
	9	0.1914	0.4306	0.5736	0.1446	0.1518	0.9115
	10	0.0960	0.3780	0.5318	0.1126	0.0670	0.6298
Swiss franc	1	0.3352	0.0158	0.1258	0.3546	0.3474	0.2674
	2	0.3222	0.2746	0.3814	0.4200	0.2270	0.9698
	3	0.1978	0.1596	0.4300	0.4282	0.5764	0.7280
	4	0.3486	0.2350	0.4332	0.5384	0.2730	0.8505
	5	0.2638	0.1590	0.4056	0.2466	0.1978	0.4710
	6	0.2866	0.1222	0.3014	0.3452	0.1690	0.4498
	7	0.2362	0.2614	0.4920	0.3482	0.1992	0.6365
	8	0.1620	0.5298	0.4080	0.4102	0.3216	0.9575
	9	0.2254	0.2530	0.3948	0.5756	0.6350	0.9999
	10	0.2878	0.2448	0.3470	0.6600	0.3316	0.9999

Notes: The table reports p-values associated with coefficient estimates derived from the Evans and Lyons (2002a) model (equation (60) above) using RBS aggregate customer order data. P-values calculated as above, but excluding the initial 194 sample observations. The F-stat is a χ^2 test of joint significance of the estimated coefficients. The number of simulations at each forecast horizon is adjusted where necessary to account for presence of collinearity. Details available on request.

5 Can An Old Lady Keep A Secret? A Microstructural Study of Policy Announcements at the Bank of England

5.1 Introduction[88]

The Bank of England (BoE), fondly known as the "Old Lady of Threadneedle Street," was granted operational independence to set its key repurchase, or 'repo', rate by the incoming Labour government in 1997 with the goal of creating policy consistent with price stability and economic growth. In practice, interest rate decisions are made by the Bank's Monetary Policy Committee (MPC), which meets for two days each month and issues a statement regarding interest rate decisions at noon on the second meeting day. This framework allows a natural laboratory setting for examining the impact of monetary policy decisions around a known time and date. Since the market knows that the interest rate announcement arrives at noon, there may be positioning prior to the announcement and news effects after the announcement that result in systematic patterns in exchange rate behaviour on MPC meeting days that differ from other days. We examine the evidence in the foreign exchange market to analyse the pattern of exchange rate changes and volatility surrounding the noon announcement.

One hypothesis to be explored is that positioning prior to the policy announcement could involve informed traders having superior information regarding the policy outcome. This need not involve information leaks of inside "secrets" from the MPC, but instead could reflect the activity of market participants adept at reading the public signals regarding the state of the economy and their interpretation of the likely MPC response to these signals. In addition, since activities directly related to each MPC meeting are spread over three different days, the analysis will include an examination of the pre-meeting briefing day, the first day of the meeting, and the second day of the meeting when the policy decision is made and publicised.

Our focus is on the response to meeting activities in the foreign exchange market, specifically the sterling-dollar exchange rate. Both daily and high-frequency, intraday data are employed in the analysis. The daily data provide a bird's eye view of the market around MPC meetings and then, given the findings from this low-frequency analysis, a microscope is taken to the data to examine exchange rate dynamics on days related to meetings. The intraday econometric framework is provided by a Markov switching model where exchange rate returns switch between a high-volatility, informed-trading state, and a low-volatility, uninformed or liquidity trading state on MPC days. A key difference from the usual Markov switching model employed in financial analysis is our incorporation of endogenous shifts in the transition probabilities where the shifts are modeled as a function of variables related to the MPC meeting and policy outcomes.

[88] This paper was written with Michael Melvin and Mark Taylor.

The next section provides institutional details on the MPC and the policy-setting process. Section 3 contains a discussion of the econometric methodology and hypotheses to be examined. Section 4 covers the data and then reports the empirical findings. Finally, Section 5 summarizes our conclusions and discusses directions for future research.

5.2 The Monetary Policy Committee

In May 1997 Gordon Brown, the Chancellor of the Exchequer, announced that the BoE would be given the responsibility for setting interest rates via the new MPC[89]. The MPC was to focus on an inflation target of 2.5 percent for the retail price index excluding mortgage interest payments[90]. Conditional on maintenance of the inflation target, the MPC could also address fluctuations in economic growth and employment.

The MPC is comprised of nine members. Five are drawn from the BoE: the Governor, the two Deputy Governors, and two Executive Directors. The other four members are drawn from outside the Bank and are appointed by the Chancellor of the Exchequer. At the time this paper was written, the four external members include two academic economists and two business economists. The Governor serves as the Committee chair.

The Committee meets monthly, normally on the Wednesday and Thursday following the first Monday of each month. The meeting dates for each year are published well in advance of the meetings. The Friday morning prior to each meeting, the Committee meets for a briefing to prepare for the meeting. Summaries of important news and trends are provided by senior BoE staff. On the Monday and Tuesday prior to the meeting, the BoE staff prepares any additional background information and analysis required by the Committee. On these days MPC members receive written answers to any questions that arose at the Friday briefing along with any new data releases or important news.

The meeting typically begins at 3pm on Wednesday afternoon with a review of the state of the economy and a discussion of key issues. The Chief Economist of the BoE starts the meeting with a short summary of any major events since the Friday briefing. Then there is open discussion among the Committee regarding the news and state of the economy.

On Thursday morning, the MPC reconvenes and the Governor begins with a summary of the major issues. Members are then invited to state their views of the appropriate policy to follow. The Deputy Governor responsible for monetary policy will usually speak first with the Governor speaking last. Ultimately, the Governor offers a motion that he suspects will result in a majority vote and then calls for a vote. Members vote with a one-member, one-vote rule. Those in the minority are asked to state their preferred level of interest rates. Lastly, the press statement is developed. If the decision is to change interest rates or follow a policy that was not expected by the market, the press statement will include

[89] For institutional background on the MPC and the monetary policy process, see Bean (1999).

[90] This target was changed in 2003 to 2.0 percent for the Harmonised Consumer Price Index.

the reasons for the action taken. In other cases, simply the decision is reported. The decision of the MPC is announced at noon, London time. Following the announcement, policy is implemented with open-market operations beginning at 12:15.

5.3 Methodology

The focus of this paper is on inference regarding shifts in the sterling-dollar exchange rate during MPC meetings. Given that the market knows when the MPC meets and when the decisions are announced, we want to examine the evidence regarding any market positioning during the meeting along with the evidence regarding the news content of the meeting. A related goal is to explore whether MPC meeting days are different from other days in terms of systematic patterns in sterling-dollar.

First, we will examine the evidence around MPC meetings. As discussed above, given the multi-day structure of MPC deliberations, one may hypothesise that the market forms an opinion about the likely meeting outcome prior to the public announcement at noon on the second day of the meeting. This does not have to rest upon information leaks from the Committee, but on astute MPC-watchers' informed opinions of the likely Committee vote. An analogy in the Federal Reserve case is the oft-cited story of how Fed-watchers at one time gauged the likely FOMC decision by the size of the briefcase that Alan Greenspan carried to work. The idea was that a thick briefcase signaled a likely interest rate shift while a thin briefcase signaled a high probability of no change in policy. No doubt, there are many such stories one could gather from MPC watchers as well.

We will explore the evidence in the data regarding briefing days, first meeting days, and second meeting days by initially analysing daily returns on the sterling-dollar exchange rate. Daily data will be employed to examine the following questions:

D1. Are exchange rate returns different on days when the MPC meets?

This question is addressed by examining characteristics of the distribution of exchange rate returns on briefing days, first meeting days, and second meeting days compared to all days. The key tests are for equality of means and variances across the different days.

D2. Is information on MPC meeting days useful in explaining daily exchange rate returns?

We estimate models of daily exchange rate returns incorporating dummy variables for days of MPC briefings, first, or second meeting days. In addition, we incorporate a variable on the size of the interest rate change to estimate models of the following form:

$$\Delta e_d = \alpha + \beta_1 Briefing + \beta_2 Day1 + \beta_3 Day2 + \beta_4 \Delta i + u_t \qquad (64)$$

where Δe_d is the change in the logarithm of the exchange rate on day d, and Briefing, Day1, and Day2 are dummy variables equal to 1 on days where an MPC briefing occurred, or for first or second days of MPC meetings, respectively, and equal to zero otherwise. The variable Δi is the change in the interest rate on days where the MPC voted to change rates.

D3. Is information on MPC meeting days useful in explaining extreme exchange rate events?

We use a sample of daily exchange rate data over the period of 1997-2002 to estimate the standard deviation of the exchange rate return. An "extreme" event is defined as a day when the absolute sterling-dollar exchange rate return exceeds 2.5 standard deviations. We then create a binary variable equal to 1 on days of extreme returns and 0 otherwise. A logit model is employed to estimate the likelihood observing an extreme event as a function of the MPC meeting activities.

Beyond the evidence in daily data regarding differences across MPC meeting days and other da‘ ys, we take a microscope to the data for second meeting days to examine the intraday evidence on days when a policy decision is announced. Before turning to the questions to be examined, the econometric framework employed in the analysis is first introduced.

Clearly, some questions call for high-frequency data. We typically think of high-frequency exchange rates on any given day trading within a fairly narrow band with first-order autocorrelation. However, on days of MPC meetings, we may expect important news to be received by the market so that the underlying data generating process delivers shifts in the exchange rate regime. One popular method of modeling nonlinear regime switches is the Markov switching model associated with Hamilton (1990, 1994). Following in this tradition, a Markov-switching first-order autoregressive model for exchange rate returns is postulated as follows:

$$\Delta e_t = \mu(S_t) + \rho(S_t)[\Delta e_{t-1} - \mu(S_t)] + \varepsilon_t \tag{65}$$

$$\varepsilon_t \sim N[0, \sigma^2(S_t)] \tag{66}$$

where Δe_t is the change in the logarithm of the exchange rate at time t. Note that the mean of the exchange rate returns process, μ, the autocorrelation coefficient, ρ, and the variance of the innovation, ε_t, are allowed to take on one of two values depending on the realisation of an unobserved state variable $S_t \in \{1, 2\}$. We assume that the state variable S_t evolves according to a two-state Markov process. One of the states (say, state 2) may be thought of as reflecting the usual pattern of exchange rate returns with zero mean and relatively small variance. This "tranquil" state is the normal state that would be associated with liquidity trading when no important new information arrives in the market. The other state (say, state 1) may be thought of as the informed-trading state when volatility is high and realized returns may be much larger than normal.

So far the methodology proposed looks familiar from applied studies such as Engel and Hamilton (1990). However, we diverge from the traditional Markov approach by modeling the probabilities of switching from one regime to another endogenously. Thus, if we denote the transition probability of switching from regime j to regime i at time t as p_t^{ij} for $i, j \in \{1, 2\}$, then we can write the postulated functions for the transition probabilities, conditional upon information at time t, I_t, and the previous state, as follows:

$$p_t^{ii} \equiv \Pr[S_t = i \mid S_{t-1} = i, I_t] \equiv \Phi[\alpha_{ii} + \beta_{ii}' X_t] \tag{67}$$

for $i \in \{1, 2\}$, where $\Phi[]$ denotes the cumulative normal density function (in order to ensure that the probabilities lie in the unit interval) and where $X_t \in I_t$ is a vector of variables known at time t which may influence the transition probability according to the vector of loadings β_i. Given p_t^{11}, we implicitly have $p_t^{21} = 1 - p_t^{11}$. Similarly, given an estimate of p_t^{22}, we implicitly have $p_t^{12} = 1 - p_t^{22}$. Estimation of the model was carried out using a modified version of the EM algorithm due to Diebold, Lee and Weinbach (1994).

The Markov-switching framework is employed to address several questions of interest in the intraday setting. First, we explore the following:

I1. Are meeting decisions "news" such that there is an exchange rate response following the noon announcement?

A variety of indicator variables were considered for elements of the explanatory variable vector. To test if the policy announcement released at noon is price-relevant public news, we incorporate various dummy variables equal to 1 for a certain afternoon period and equal to 0 otherwise. We experiment with alternative time dummies following noon as a sensitivity analysis.

A second question of interest is:

I2. Is there evidence of positioning during the meetings prior to the policy announcement at noon?

To address this question, we incorporate dummy variables that equal 1 for various intervals of time prior to noon and equal 0 otherwise. We explore alternative definitions over different morning time intervals as a sensitivity analysis.

5.4 Data and Empirical Findings

The analysis of the foreign exchange market on MPC meeting days involves both daily and high-frequency, intraday data on the sterling-dollar exchange rate. Table 1 lists the MPC meeting days in our sample and the decision taken. There were 65 meetings from the first, in June 1997, through to the end of our sample in September, 2002. Interest rates were raised at 9 meetings and lowered at 14 meetings. The presentation of results is organized by reviewing the daily data and empirical results first followed by the intraday analysis.

5.4.1 Daily Data and Results

Daily observations on the sterling-dollar exchange rate were obtained from the Federal Reserve Board. These are buying rates at noon New York time (17:00 London time). The daily data are sampled for the period May 1, 1997 to September 30, 2002. Descriptive statistics for the level of the exchange rate and the first difference of the log of the exchange rate (the returns) are presented in Table 2. The table shows the mean value of the exchange rate over the sample period as USD1.5572 with a high of USD1.7222 and a low of USD1.3730. This was a period with considerable exchange rate volatility. The summary statistics in Table 2 suggest that the exchange rate distributions are non-normal. In addition, the last row reports the Augmented Dickey-Fuller (ADF) test statistic for a unit root. One cannot reject the nonstationarity of the level of the exchange rate but exchange rate returns are found to be stationary. In the empirical work below, exchange rate returns are utilized.

Turning to the questions raised in the previous section, the first issue to be addressed with the daily returns is:

D1. Are exchange rate returns different on days when the MPC meets?

Tests for equality of means and variances of MPC meeting days versus other days are reported in Table 3. The table compares all second meeting days, first meeting days, and briefing days with all other days. In addition, results are also reported for those meeting days associated with an interest rate change.

The test for equality of means is based on a single-factor, between-groups, analysis of variance (ANOVA). If two groups of days have the same mean, then the sample means between groups should have the same variability as any within group mean. The table reports the p-value associated with each t-statistic (square root of the F-statistic) for each paired group of days. For instance, the first row reports the test of the equality of means between all second MPC meeting days (when decisions are announced) with all other days. This row has the smallest p-value of 0.27, but none of the differences in means for grouped pairs of days are statistically significant.

The equality of variances is based on the F-statistic associated with the variance ratio for the paired groups of days. For example, the fourth row reports a p-value of 0.15 for all second meeting days when an interest rate change was announced versus all other days. The value of 0.15 is the smallest in the table, so the evidence suggests that there are no statistically significant differences in variances across the various paired groups of days.

Taken as a whole, the evidence in Table 3 indicates that one cannot distinguish MPC meeting days from other days on the basis of means and variances of daily exchange rate returns. This then brings us to the next question:

D2. Is information on MPC meeting days useful in explaining daily exchange rate returns?

To answer this question, we first estimated the model represented by equation (65) above by OLS. The evidence indicates that the explanatory variables $Briefing$, $Day1$, $Day2$, and Δi had no power in explaining exchange returns. This was true whether the meeting-related variables represented all meeting days or just those with interest rate changes. However, the OLS estimates did indicate that significant GARCH effects were present. Estimating a GARCH specification for daily exchange rate returns and incorporating the dummy variables related to MPC meetings in the variance equation indicated no explanatory power for variables related to all meetings. However, using variables associated with meetings where interest rates were changed generated the values reported in Table 4. In this specification, the dummy for meeting day 2, $Day2$, and the absolute value of the announced change in the interest rate, $abs(\Delta i)$, both have statistical significance with p-values of 0.05 and 0.04 respectively. However, the overall statistical significance of the four meeting-related variables is statistically insignificant with a likelihood-ratio statistic of 3.378 and associated p-value of 0.50.

To summarize the findings on exchange rate returns, the returns are well characterised by mean zero changes and the meeting day information has no explanatory power for returns. However, there is some evidence of greater volatility on meeting day 2 when interest rate changes are announced. Then, given that there is an announced interest rate change on day 2, the greater is the absolute value of the interest rate change, the lower the volatility. So days when interest rate changes are announced appear to be different from other days. Note that the unconditional volatility tests reported in Table 3 indicated that second meeting days with interest rate changes had the lowest p-value among the tests of equality of variance reported. Yet the p-value of 0.15 indicated a lack of statistical significance. Now, with a well-specified conditional volatility model, we find evidence of second meeting days with interest rate changes to have a significantly different conditional variance than other days.

The final question addressed to the daily returns is:

D3. Is information on MPC meeting days useful in explaining extreme exchange rate events?

We define an "extreme" exchange rate event as a day with a greater than 2.5 standard deviation return. The standard deviation of returns calculated over all days is shown in Table 2 as being equal to 0.004823. So an extreme exchange rate return day is one where the absolute value of the return exceeds 0.01206, or a change in the exchange rate in excess of 1.2 percent. There were 28 days, out of the 1,363 days in the sample, with such a return.

To answer the question of whether MPC meeting–related information is related to extreme exchange rate returns, a binary dependent variable is created equal to 1 on days when the exchange rate return exceeded 0.01206 in absolute value and equal to 0 otherwise. Then a Logit model is employed to estimate the probability of an extreme exchange rate change as a function of the MPC-related variables. Only second meeting days when an interest rate change occurs are significantly related to extreme exchange rate returns. Table 5 reports estimation results. The dummy variable for second meeting days, *Day*2, enters the regression with a coefficient of 1.5714 and a p-value of 0.04. Through simulation methods utilizing the estimates reported in Table 5, the probability of observing an extreme exchange rate return is estimated to increase by a factor of 0.3928. So there is about a 40 percent increase in the probability of observing an extreme exchange rate event on the second day of MPC meetings when an interest rate change occurs.

5.4.2 Intraday Data and Results

Indicative quotes on the sterling-dollar exchange rate were obtained from HSBC. Since the daily results indicate that only the second MPC meeting days appear to be different from other days, we now take a microscopic look at these days.

In the high-frequency setting, our references to MPC meeting days is always with regard to second meeting days when the policy votes are taken. For each MPC meeting day we identified a control day as the same day of the week one week after the meeting. Then tick data on the sterling-dollar exchange rate were gathered for each meeting and control day from the first MPC meeting day in June 1997 to the September 2002 meeting. There are no data for the meeting on December 9, 1999 or for the extraordinary unscheduled meeting of September 18, 2001.

We sampled the last quote of each 5-minute interval to create a series of exchange rate returns. The returns employed in the empirical work are the change in the logarithm of the 5-minute observations multiplied by 10,000 over the hours 7:00-17:00 London time. We have 63 MPC meeting days and 63 control days. The data for each day are stacked in serial order to create a data set with 15,246 observations.

Before turning to the specific questions addressed to the high-frequency data, we first discuss some general issues related to the model and estimation. The Markov model represented by equations (66), (67), and (68) above was used to estimate the effect of MPC-related activity on the transition probabilities. Preliminary estimates indicated that the mean is zero in both states, so states were identified by variance shifts. State 1 is the high-variance state associated with information-based trading and state 2 is the low variance state associated with the normal market conditions of liquidity trading. The estimated state 1 variance is generally found to be about 2.5 times that of state 2. In terms of the transition probabilities, p^{11} is the probability of remaining in the high-volatility state and p^{22} is the probability of remaining in the low-volatility state. Normally, we expect $p^{22} \succ p^{11}$ and this is what the data reveal. Estimating a Markov-switching model with fixed transition probabilities resulted in the following estimates: $p^{11} = 0.810$ and $p^{22} = 0.945$. The unconditional probability of being in state 2 associated with these transition probabilities is given as

$$\frac{1 - p^{11}}{(1 - p^{22}) + (1 - p^{11})} = 0.78 \tag{68}$$

so the unconditional probability of being in state 1 is 0.22.

Statistically significant negative first-order autocorrelation was found in all models. Negative autocorrelation is a common finding in high frequency exchange rate returns. Rather than report all model coefficients, the results associated with the questions of interest will summarize findings regarding the key coefficients and implied transition probabilities.

Table 6 part A reports estimates of the constant transition probability model and then in part B the preferred model is presented. The payoff from estimating the endogenous transition probabilities is demonstrated by the likelihood ratio statistic of 103.88 (p-value of 0.00) associated with comparing part A as the restricted estimate and part B as the unrestricted. Transition probabilities are modeled as varying with dummy variables that switch to 1 at certain times of day and are equal to 0 otherwise. Preliminary estimates suggested that the preferred model has p^{11} a function of a constant and a dummy that is equal to one from 12:00-13:30 only on MPC meeting days when interest rates changed ($Dum_{\Delta i,12-13:30}$), while p^{22} is a function of a constant, a dummy equal to one on all MPC meeting days from 12:00-13:30 ($Dum_{\Delta i+no\Delta i,12-13:30}$), a dummy equal to one on all days from 13:30-17:00 ($Dum_{all,13:30-17}$) , a dummy equal to one from 11:30-11:55 on all MPC days ($Dum_{\Delta i+no\Delta i,11:30-11:55}$), and a dummy equal to one at noon on MPC days ($Dum_{\Delta i+no\Delta i,12}$). Estimates are reported in Table 6, part B. It is seen that each of the determinants of p^{11} and p^{22} differ significantly from zero with p-values of 0.00.

Turning now to the questions addressed with the intraday data, the first question is:

I1. Are meeting decisions "news" so that there is an exchange rate response following the noon announcement?

The results in Table 6, part B indicate that the probability of remaining in the informed trading state p^{11} is significantly higher from 12:00-13:30 following news that the MPC has raised interest rates. The implied change in p^{11} is from 0.764 before noon on days when the MPC raises interest rates to 0.933 from 12:00-13:30 on those days. The probability of remaining in the tranquil state p^{22} is significantly lower at noon on MPC meeting days. The implied probability changes from 0.954 before 11:30, to 0.906 from 11:30-11:55, and then to 0.291 at noon. In addition, p^{22} is significantly lower in late afternoon from 13:30-17:00 on all days, but one may argue that the implied change in probability from 0.954 before 13:30 to 0.927 after 13:30 is not economically significant.

Following the baseline model estimates, Table 6 reports results for the sensitivity of the transition probability models over alternative specifications using afternoon dummy variables in part C. In each case, the baseline model is augmented by an additional explanatory variable. These additional dummy variables represent the same time of day but over different days than those incorporated in the preferred specification in part B. For instance, the first row reports the results of adding ($Dum_{\Delta i+no\Delta i,12-13:30}$) to the p^{11} equation. The additional variable has no explanatory power as seen by the p-value associated with the coefficient of the additional variable of 0.37. While not shown in the table, no other results are changed in any substantive way. The next five rows of the table report results for adding other explanatory variables to the preferred specification. In no case, does any variable enter significantly. The last two rows of part C report the results of adding means to the Markov-switching specification. It is clear that the estimated means for both state 1 and state 2 do not differ significantly from zero. No other results of the preferred specification are changed by the addition of the variables in part C of Table 6.

The evidence in Table 6 presents a robust result: There is a systematic regime switch to the high-volatility informed trading state on MPC days when interest rates are changed. This effect is highly significant for the 1.5 hours following the interest rate announcement. After this time, the probability of remaining in the informed trading state falls significantly. This result for MPC days with interest rate news is clearly distinguished from other days and is not simply a "time of day" effect that exists in the market every day. The question posed above, Are meeting decisions "news" so that there is an exchange rate response following the noon announcement? warrants a strong affirmative response.

The second question addressed by the intraday evidence is:

I2. Is there evidence of positioning during the meetings prior to the policy announcement at noon?

The previous results summarised in Table 6 established that the noon announcement of interest rate changes were indeed price-relevant news as there is a switch to the high-volatility informed trading state immediately after the announcement. The current question requires that the pre-noon period receive a microscopic examination. Table 7 repeats the relevant coefficient estimates for the preferred model as initially shown in Table 6. The news anticipation effect is captured by the coefficient on the dummy variable for 11:30-11:55 on all MPC days. This variable is found to enter significantly in the p^{22} equation. The coefficient of –0.37 with a p-value of 0.00 indicates a significant drop in the probability of remaining in the low-volatility state prior to noon. So there is some evidence that the market anticipates the news and takes positions accordingly.

Parts A, B and C of Table 7 incorporate alternative dummy variables into the preferred model as a robustness check. This proceeds much like the analysis associated with the post-noon announcement effect. Starting with the baseline preferred model, we specify alternative dummy variables for the pre-noon period for our three different types of days: all days, all MPC meeting days, and MPC meeting days when an interest rate change was announced and examine the sensitivity of the estimates to the additional variables. Part A includes dummy variables for all days over alternative times of the morning. For instance, the first row of part A includes a dummy equal to 1 from 11:45-11:55 in the p^{11} equation. The p-value indicates that this additional variable has no significant explanatory power. While not reported, no other results from the preferred model are affected by the inclusion of this variable. Similarly, the other variables added to the p^{11} and p^{22} equations have no significant explanatory power except for one case. The only moderately significant variable is the dummy for 11:30-11:55 in the p^{22} equation, with a coefficient of 0.67 and a p-value of 0.07. This suggests that the probability of remaining in the tranquil state p^{22} rises for all days during this time. So the negative effect found in the preferred model for MPC days is not capturing some daily event. On non-MPC days, p^{22}rises rather than falls between 11:30-11:55.

Part B of Table 7 incorporates additional morning dummy variables for all MPC days into the preferred model. In no case does the addition of any other morning dummy add any significant explanatory power to the model. Finally, part C incorporates additional morning dummy variables for MPC days with an interest rate change into the preferred model. There is one marginally significant variable in this case. The dummy variable for 11:00-11:55 enters significantly in the p^{11} equation with a coefficient of -0.69 with a p-value of 0.08. This suggests a drop in the probability of remaining in the high-volatility state prior to the news on MPC days when interest rates change.

Taken as a whole, there is some, albeit limited, evidence of regime switching in terms of exchange rate volatility in the morning prior to the end of the MPC meetings. The evidence is strongest for the p^{22} equation for the 11:30-11:55 time period. During this interval, there is a statistically significant drop in the probability of remaining in the low-volatility state. However, the implied change in probability is from 0.954 to 0.906. So one may argue that this is not economically significant. Of course, since the meetings always end prior to the noon announcement and the MPC's policy decision is known by insiders, the regime switching could be a result of signals read by market participants. This is not to claim that there are deliberate information leaks emanating from the committee. It may be something much more subtle (remember the Greenspan briefcase story presented earlier). The evidence presented here indicates no particularly large probability shifts prior to meeting end. This is certainly true if one considers the probabilities of regime switching in the morning compared with the afternoon. The news impact appears to be much larger than any anticipation effect.

The implications of the intraday estimation results for the transition probabilities are summarized in the figure below. This plots p^{11}, the probability of remaining in the high-volatility, informed-trading state, for the three types of days in our sample as generated by the preferred model reported in Table 6. This probability is averaged across all observations for each type of day for each 5-minute interval. One can see dramatic differences in p^{11} across types of days and time of day. It is clear that non-MPC meeting days are characterized by low-volatility, liquidity trading as the probability of remaining in the informed trading state is quite low all through the day, fluctuating between 0.1 and 0.3. On MPC meeting days when no interest rate change occurs, there is an increase in the average p^{11} that begins around 11:30 and continues until noon when it peaks at about 0.65. After this peak at noon, the probability quickly falls to about 0.30 by 12:30 and then by 13:30 is quite similar to the afternoon pattern on non-MPC days. On MPC meeting days when an interest rate change occurs, there is a dramatic jump from less than 0.20 to more than 0.90 at noon when the policy announcement is released. Then p^{11} remains above 0.70 until about 13:00 after which it continues to fall so that by about 13:30 it appears to follow a pattern much like other days. The evidence in the figure indicates that MPC days are indeed different from other days. Whether interest rates change or not, the noon policy announcement appears to be price-relevant news. There is some modest evidence of positioning in advance of the announcement on MPC days when no interest rate change occurs, but for days when interest rates are changed, it appears that the market response comes immediately at noon with the news.

The impact of MPC policy announcements upon the probability of being in either volatility state may depend significantly upon whether announced policy actions are expected by market participants. Market expectations can be measured in three ways. First, news wire services such as Bloomberg and Reuters canvass the views of market economists ahead of each MPC meeting, providing a snapshot of expectations amongst approximately twenty five well-informed market participants. Second, changes in deposit interest rates between the days immediately before and after the MPC announcement day can provide evidence of any surprises in the money markets. And third, evidence of policy surprises can also be gauged from price action in the futures market, for instance the short sterling interest rate contract traded on LIFFE.[91] A policy surprise common to all three of these participant groups may be particularly relevant for the behaviour of sterling-dollar around the time of MPC policy announcements. There have been a number of common shocks since the formation of the MPC in June 1997, particularly during the initial period of the MPC's existence as market participants gained some understanding of the committee's policy reaction function. More recently, following this learning period,.there have been few common shocks using our preferred metrics. For this reason, we chose not to incorporate the impact of policy surprises into our intraday analysis of the MPC.

5.5 Summary and Conclusion

The MPC was created in 1997 to foster monetary policy consistent with stable inflation and economic growth. Since the MPC meets at regularly scheduled, pre-announced times and the policy decision is always announced at noon, the meetings provide a natural laboratory for examining exchange rate dynamics on days when monetary policy is formulated. Our particular interest is with respect to the news content of the policy announcement and whether there is any evidence of positioning during the meeting prior to the announcement.

Daily data on the sterling-dollar exchange rate are employed to analyse any differences that may exist regarding exchange rate returns on the three kinds of days associated with MPC meetings: the pre-meeting briefing day; the first day of the meeting; and the second day of the meeting when the policy announcement is made. First, we examine tests of the equality of means and variances for the three different types of days versus all other days. There is no evidence of any statistically significant difference in means or variances across days. However, the test for the equality of the variance on second meeting days versus all other days had the lowest p-value of 0.15. Second, we estimate models of daily exchange rate returns to infer if information on MPC meeting days contains any explanatory power. Estimation results suggest that daily exchange rate returns are well characterized by mean zero changes and meeting day information has no explanatory power for returns. However, evidence of strong GARCH effects in the daily returns was found and incorporation of MPC meeting day information in the conditional variance equation revealed evidence of greater conditional volatility on second meeting days when interest rates are changed. Third, we estimate the probability of observing an extreme exchange rate event, defined as a return in excess of 2.5 standard deviations of the mean return, as a function of information related to MPC meetings. The evidence suggests that the probability of observing an extreme exchange rate change increases by about 40 percent on the second day of MPC meetings when an interest rate change is

[91] London International Financial Futures Exchange.

announced.

The evidence from daily data suggests that only the second days of MPC meetings are different from other days in terms of exchange rate dynamics. After the bird's-eye view given by the daily data, we then turn to a microscopic view of second meeting days provided by intraday exchange rate returns. A high-frequency sample of 5-minute observations over 7:00-17:00 London time is analysed using a Markov-switching framework. We assume that there exist two states: state 1, the high-volatility state associated with informed trading, and state 2, the low-volatility state associated with liquidity trading. We diverge from the usual non-linear regime-switching framework to model endogenous transition probabilities as a function of information regarding the meeting days. The transition probabilities are found to systematically switch on meeting days. The probability of remaining in the high volatility state is estimated to increase from 0.764 before noon to 0.933 from 12:00-13:30 on MPC days when interest rates are changed. In addition, the probability of remaining in the low volatility state is estimated to fall from 0.954 before 11:30, to 0.906 from 11:30-11:55, to 0.291 at noon on MPC meeting days. So the evidence indicates that there is a statistically and economically significant news effect related to the noon announcement.

Regarding anticipation of the policy announcement, the finding that the probability of remaining in the low-volatility state falls from 0.954 before 11:30 to 0.906 from 11:30-11:55 is statistically significant, but is of very limited economic significance. This is the strongest evidence found for any policy anticipation effect of market positioning in anticipation of the announcement. So to answer the question posed in the title: Can an old lady keep a secret? The answer appears to be yes. The second day of MPC meetings is best characterized as having a strong exchange rate reaction to the news announcement at noon with little evidence of positioning during the morning period of the meeting.

References

[1] Bean, C. (1999), The formulation of monetary policy at the Bank of England. Bank of England Quarterly Bulletin, May, 434-441.

[2] Danielsson, J. and R. Payne (2002), Real Trading Patterns and Prices in Spot Foreign Exchange Markets. Journal of International Money and Finance 21 (2), 203-222.

[3] Diebold, F., J.H. Lee and G. Weinbach (1994), Regime Switching with Time-Varying Transition Probabilities. In: Hargreaves, C. (Ed.), Nonstationary Time Series Analysis and Cointegration. Oxford University Press, Oxford and New York, pp. 283-302.

[4] Engel, C and J.D. Hamilton (1990), Long Swings in the Dollar: are They in the Data and Do Markets Know it? American Economic Review 80 (4), 689-713.

[5] Hamilton, J.D. (1990), Analysis of Time Series Subject to Changes in Regime. Journal of Econometrics 45, 39-70.

[6] Hamilton, J.D. (1994), Time Series Analysis. Princeton University Press, Princeton, NJ.

Table 1: Monetary Policy Committee Meetings and Interest Rate Decisions

DATE	INTEREST RATE DECISION
June 6, 1997	Up 1/4 %
July 10, 1997	Up 1/4 %
August 7, 1997	Up 1/4 %
September 11, 1997	No change
October 9, 1997	No change
November 6, 1997	Up 1/4 %
December 4, 1997	No change
January 8, 1998	No change
February 5, 1998	No change
March 5, 1998	No change
April 9, 1998	No change
May 7, 1998	No change
June 4, 1998	Up 1/4 %
July 9, 1998	No change
August 6, 1998	No change
September 10, 1998	No change
October 8, 1998	Down 1/4 %
November 5, 1998	Down 1/2 %
December 10, 1998	Down 1/2 %
January 7, 1999	Down 1/4 %
February 4, 1999	Down 1/2 %
March 3, 1999	No change
April 8, 1999	Down 1/4 %
May 6, 1999	No change
June 10, 1999	Down 1/4 %
July 8, 1999	No change
August 5, 1999	No change
September 8, 1999	Up 1/4 %
October 7, 1999	No change
November 4, 1999	Up 1/4%
December 9, 1999 (Data unavailable)	No change

Table 1 (cont.): Monetary Policy Committee Meetings and Interest Rate Decisions

DATE	INTEREST RATE DECISION
January 13, 2000	Up 1/4 %
February 10, 2000	Up 1/4 %
March 9, 2000	No change
April 6, 2000	No change
May 4, 2000	No change
June 7, 2000	No change
July 6, 2000	No change
August 3, 2000	No change
September 7, 2000	No change
October 5, 2000	No change
November 9, 2000	No change
December 7, 2000	No change
January 11, 2001	No change
February 8, 2001	Down 1/4 %
March 8, 2001	No change
April 5, 2001	Down 1/4 %
May 10, 2001	Down 1/4 %
June 6, 2001	No change
July 5, 2001	No change
August 2, 2001	Down 1/4 %
September 6, 2001	No change
September 18, 2001 (data unavailable)	Down 1/4%
October 4, 2001	Down 1/4 %
November 8, 2001	Down 1/4%
December 5, 2001	No change
January 10, 2002	No change
February 7, 2002	No change
March 7, 2002	No change
April 4, 2002	No change
May 9, 2002	No change
June 6, 2002	No change
July 4, 2002	No change
August 1, 2002	No change
September 5, 2002	No change

Table 2: Descriptive Statistics for sterling-dollar Exchange Rate

	Exchange Rate	d(log(Exchange Rate)
Mean	1.5572	−0.000023
Median	1.5890	−0.000129
Maximum	1.7222	0.019967
Minimum	1.3730	−0.020414
Standard Deviation	0.0936	0.004823
Skewness	−0.2530	0.11245
Kurtosis	1.5323	4.0908
Jarque-Bera Statistic*	0.00	0.00
ADF test statistic	−1.5285	−16.3970

Notes: The table reports daily exchange rate data for the sterling-dollar exchange rate over the period of May 1, 1997 to September 30, 2002. Descriptive statistics for both the exchange rate level and returns are reported. * indicates p-value.

Table 3: Tests for Equality of Exchange Rate Return Means and Variances: MPC Meeting Days versus Other Days

Variable	Mean test	Variance test
All 2^{nd} meeting days	0.27	0.38
All 1^{st} meeting days	0.97	0.34
All briefing days	0.43	0.25
2^{nd} meeting days with interest rate change	0.95	0.15
1^{st} meeting days before interest rate change	0.70	0.69
Briefing days before interest rate change	0.76	0.66

Notes: The table reports p-values for tests of the equality of means and variances between the MPC event days listed in the first column and the rest of the sample days. The mean equality test is a t-test and the equality of variances is tested with an F-statistic. The 2nd meeting days are the days when the policy decision is announced at noon (typically a Thursday). The 1st meeting days are the days when the afternoon discussions are held (typically a Wednesday). The briefing days refer to the Friday information sessions that are held at the end of the week prior to the MPC meetings. The last 3 rows of the table reports the p-values for MPC meetings where interest rates were changed.

Table 4: GARCH Model of Daily Exchange Rate Returns

Variable	Coefficient Estimate	p-value
Mean equation:		
Constant	000012	0.92
AR(1)	0.0398	0.15
AR(2)	−0.0541	0.08
Variance equation:		
Constant	$5.98x10^{-7}$	0.02
ARCH(1)	0.0269	0.00
ARCH(2)	0.9461	0.00
Briefing	$-1.02x10^{-6}$	0.78
Day1	$-3.10x10^{-6}$	0.56
Day2	$1.42x10^{-5}$	0.05
abs(Δi)	$-3.02x10^{-5}$	0.04
log-likelihood	5354.87	
Q(12)	225	

Notes: The dependent variable in the mean equation is daily sterling-dollar exchange rate returns over the period May 1, 1997 to September 30, 2002. Dummy variables related to MPC meetings when interest rates were changed are incorporated in the variance equation. Dummies equal 1 on the day specified and zero otherwise. In addition, the absolute value of the announced interest rate change is included in the variance equation.

Table 5: Logit Model of Extreme Exchange Rate Returns

Variable	Coefficient Estimate	p-value
Constant	−3.9227	0.00
Day2	1.5714	0.04
log-likelihood	−135.04	

Extreme exchange rate returns are defined as those exceeding 2.5 standard deviations based upon the sample standard deviation of returns over the sample period of May 1, 1997 to September 30, 2002. A binary dependent variable equal to 1 on days with extreme returns and 0 otherwise is used in the estimation. Only the dummy variable for the second day of MPC meetings when an interest rate change occurred was found to have a significant relationship.

Table 6: Markov-Switching Model of MPC News and News Anticipation Effects

A. Constant Transition Probability Model

	p^{11}	p^{22}
Constant	0.88	1.60
	(0.00)	(0.00)
Log Likelihood	−44302.72	

B. Preferred Model

	p^{11}	p^{22}
Constant	0.72	1.68
	(0.00)	(0.00)
$dum_{\Delta i,12-13.30}$	0.78	
	(0.00)	
$dum_{\Delta i+no\Delta i,12-13.30}$		−0.34
		(0.00)
$dum_{all,13.30-17}$		−0.22
		(0.00)
$dum_{\Delta i+no\Delta i,11.30-11.55}$		−0.37
		(0.00)
$dum_{\Delta i+no\Delta i,12}$		−1.90
		(0.00)
Log Likelihood	−44250.78	

The table reports estimates of a Markov-switching model for sterling-dollar exchange rate returns sampled at a frequency of 5-minutes over the London business day. Transition probabilities are modeled as switching endogenously as a function of MPC-related events as in (ϕ denotes the cumulative normal density function):

$$p^{11} = \Phi(\alpha_{11} + \sum_k \beta_{11,k} dum_k) \tag{69}$$

and

$$p^{22} = \Phi(\alpha_{22} + \sum_k \beta_{22,k} dum_k) \tag{70}$$

Dummy variables are equal to 1 for time-of-day indicated and for days identified by the following notation: Δ_i, MPC days when interest rates were changed; $\Delta_i + no\Delta_i$, all MPC days; and all_i, all days. P-values in parentheses in table.

Table 6 (cont.): Markov-Switching Model of MPC News and News Anticipation Effects

C. Robustness Check over Different Days

	Coefficient	*Logl*
p^{11} & $Dum_{\Delta i+no\Delta i,12-13.30}$	0.17	−44250.37
	(0.37)	
p^{11} & $Dum_{all,12-13.30}$	0.14	−44250.32
	(0.33)	
p^{22} & $Dum_{\Delta i,12-13.30}$	0.40	−44249.91
	(0.20)	
p^{22} & $Dum_{all,12-13.30}$	0.06	−44250.61
	(0.55)	
p^{22} & $Dum_{\Delta i+no\Delta i,13.30-17}$	0.06	−44250.59
	(0.52)	
p^{22} & $Dum_{\Delta i,13.30-17}$	−0.11	−44250.36
	(0.37)	
μ_1	0.14	−44250.30
	(0.33)	
μ_2	−0.00	
	(0.93)	

Notes: Table shows result of inclusion of additional variables in the preferred model specification of Part B (same times as in B but for different days). P-values in parentheses in table.

Table 7: Markov-Switching Model of MPC News Anticipation Effects

A. Alternative specifications to Preferred Model reported in Table 6

	$Coefficient$	$Logl$
p^{11} & $Dum_{all,11.45-11.55}$	0.48 (0.44)	−44250.32
p^{11} & $Dum_{all,11.30-11.55}$	0.15 (0.59)	−44250.64
p^{11} & $Dum_{all,11.15-11.55}$	−0.07 (0.75)	−44250.73
p^{11} & $Dum_{all,11.00-11.55}$	−0.07 (0.67)	−44250.70
p^{11} & $Dum_{all,9.00-11.55}$	0.08 (0.41)	−44250.48
p^{22} & $Dum_{all,11.45-11.55}$	0.06 (0.77)	−44.250.74
p^{22} & $Dum_{all,11.30-11.55}$	0.67 (0.07)	−44248.09
p^{22} & $Dum_{all,11.15-11.55}$	0.62 (0.13)	−44246.72
p^{22} & $Dum_{all,11.00-11.55}$	0.21 (0.15)	−44249.66
p^{22} & $Dum_{all,9.00-11.55}$	−0.03 (0.69)	−44250.71

Table 7 (cont.): Markov-Switching Model of MPC News Anticipation Effects

B. Alternative specifications to Preferred Model

	Coefficient	*Logl*
p^{11} & $Dum_{\Delta i+no\Delta i,11.45-11.55}$	0.84	−44250.11
	(0.47)	
p^{11} & $Dum_{\Delta i+no\Delta i,11.30-11.55}$	1.67	−44248.49
	(0.62)	
p^{11} & $Dum_{\Delta i+no\Delta i,11.15-11.55}$	0.09	−44250.75
	(0.83)	
p^{11} & $Dum_{\Delta i+no\Delta i,11.00-11.55}$	−0.16	−44250.56
	(0.50)	
p^{11} & $Dum_{\Delta i+no\Delta i,9.00-11.55}$	−0.04	−44250.73
	(0.73)	
p^{22} & $Dum_{\Delta i+no\Delta i,11.45-11.55}$	5.42	−44248.02
	(0.99)	
p^{22} & $Dum_{\Delta i+no\Delta i,11.30-11.55}$	−0.00	−44250.78
	(0.99)	
p^{22} & $Dum_{\Delta i+no\Delta i,11.15-11.55}$	0.52	−44250.23
	(0.47)	
p^{22} & $Dum_{\Delta i+no\Delta i,11.00-11.55}$	−0.07	−44250.73
	(0.74)	
p^{22} & $Dum_{\Delta i+no\Delta i,9.00-11.55}$	−0.09	−44250.27
	(0.28)	

Table 7 (cont.): Markov-Switching Model of MPC News Anticipation Effects

C. Alternative specifications to Preferred Model

	Coefficient	*Logl*
p^{11} & $Dum_{\Delta i,11.45-11.55}$	−1.26 (0.13)	−44249.73
p^{11} & $Dum_{\Delta i,11.30-11.55}$	−0.73 (0.12)	−44250.00
p^{11} & $Dum_{\Delta i,11.15-11.55}$	0.08 (0.84)	−44248.42
p^{11} & $Dum_{\Delta i,11.00-11.55}$	−0.69 (0.08)	−44249.21
p^{11} & $Dum_{\Delta i,9.00-11.55}$	0.01 (0.96)	−44250.78
p^{22} & $Dum_{\Delta i,11.45-11.55}$	0.53 (0.55)	−44249.09
p^{22} & $Dum_{\Delta i,11.30-11.55}$	0.06 (0.95)	−44248.39
p^{22} & $Dum_{\Delta i,11.15-11.55}$	0.77 (0.99)	−44247.98
p^{22} & $Dum_{\Delta i,11.00-11.55}$	0.90 (0.37)	−44249.19
p^{22} & $Dum_{\Delta i,9.00-11.55}$	0.02 (0.86)	−44250.77

Figure 2: Probability of Informed Trading State