Order Flows and The Exchange Rate Disconnect Puzzle

First Version: June 2008

Martin D. D. Evans[1]
Department of Economics and McDonough School of Business
Georgetown University, Washington DC 20057;
and the National Bureau of Economic Research
(202) 687-1570: evansm1@georgetown.edu

Abstract

The aim of this paper is to establish the link between the high frequency dynamics of spot exchange rates and developments in the macroeconomy. To do so, I first present a theoretical model of exchange-rate determination that bridges the gap between existing microstructure and traditional models. I then report empirical evidence that strongly supports the presence of the link between the macroeconomy, order flow, and high frequency exchange rate returns implied by the model. In fact, my empirical results indicate that between 20 and 30 percent of the variance in excess currency returns over one- and two-month horizons can be linked back to developments in the macroeconomy. This level of explanatory power is an order of magnitude higher than that found in traditional models – even the newly developed monetary models incorporating central banks reaction functions. Moreover, it provides a straightforward solution to the exchange-rate disconnect puzzle. Namely, the high frequency behavior of spot exchange rates reflects the flow of new information reaching dealers concerning the slowly evolving state of the macroeconomy, rather than the effects of shocks that drive rapidly changing macroeconomic conditions.

Keywords: Exchange Rate Dynamics, The Exchange Rate Disconnect Puzzle, Microstructure, Order Flow.

JEL Codes: F3, F4, G1

Introduction

In the 25 years since Meese and Rogoff published their classic paper, Meese and Rogoff (1983), little progress has been made in developing models that can consistently and reliably relate the behavior of spot exchange rates to developments in the macroeconomy at frequencies higher than a year. Indeed, the lack of empirical success for standard exchange-rate models based on macro variables has given credence to the view that exchange rates are largely disconnected from macroeconomic fundamentals; a view Obstfeld and Rogoff (2001) call the exchange-rate disconnect puzzle. This is not to say that exchange rates appear disconnected from everything. In the past decade a new literature has emerged documenting a strong link between spot rate dynamics and order flows – the transaction flows arising from trades between counterparties in the foreign exchange market. In contrast to the empirical literature on standard exchange-rate models, this new microstructure literature shows that order flows account for a large fraction of the variability in spot rates over horizons ranging from a few minutes to a couple of months. For example, Evans and Lyons (2002a & 2002b) found that order flows account for between 44 and 78 percent of the daily variation in the spot exchange rates of major currency pairs. Subsequent research has established that this strong relationship between order flows and spot rates holds across different currencies, measures of order flow and trading mechanisms (see Osler 2008 and Sager and Taylor 2007 for recent surveys).

The contrasting results from the traditional and microstructure literatures have made little impact on the way researchers formulate exchange-rate models. For example, recent research on monetary models by Mark (2005), Engel and West (2006) and Engel et al. (2007) stress that short-run movements in spot rates are primarily determined by changes in expectations, principally expectations concerning the future course of interest rates controlled by central banks. These models have some empirical success at long horizons (see below), but they are completely silent on why order flow appears to be so important at short horizons. The implicit belief in this line of research seems to be that microstructure factors linking order flow and exchange rates are unrelated to the macroeconomic factors included in the model. In a similar vein, microstructure exchange-rate models, such as Evans and Lyons (1999), are silent on the potential links between the macroeconomy and order flows. These models focus on the details of currency trading between different market participants rather than the macroeconomic factors that ultimately drive the individual trades. Consequently, they cannot explain why the link between the macroeconomy and spot rate is only discernible at long horizons. Nor do they provide any clean implications for how exchange rates respond to macroeconomic policies. In short, an arm's-length observer of the literature might well conclude that the apparent disconnect between spot rates and the macroeconomy is matched by the disconnect between the traditional and microstructure approaches to exchange-rate modelling.

The aim of this paper is to establish the link between the high frequency dynamics of spot exchange rates and developments in the macroeconomy. To do so, I first present a theoretical model of exchange-rate determination that bridges the gap between existing microstructure and traditional models. I then report empirical evidence that strongly supports the presence of the link between the macroeconomy, order flow, and high frequency exchange rate returns implied by the model. In fact, my empirical results indicate that between 20 and 30 percent of the variance in excess currency returns over one- and two-month horizons can be linked back to developments in the macroeconomy. This level of explanatory power is an order of magnitude higher than that found in traditional models – even the newly developed monetary models incorporating central banks reaction functions. For example, Mark (2005) and Engel and West (2006) find

that the correlation between the log level of the real exchange rate implied by their models and the actual rate is approximately 0.3, but this encouraging result does not carry over to *changes* in log spot rates (i.e., depreciation rates). In particular, their results imply that macroeconomic factors account for just one percent of the variance in the currency depreciation rates at monthly and quarterly horizons.

The empirical success of my micro-based approach rests on three key elements. First, the model recognizes the fact that spot exchange rates are literally determined as the foreign currency prices quoted by dealers in the foreign exchange market. Second, only a limited amount of information is available concerning the *current* state of the macroeconomy to dealers and other agents when they trade foreign currency. These two elements allow me to theoretically identify the role order flow plays in linking spot rates to the macroeconomy. In particular, the model shows that the order flow generated by trades between dealers and agents can convey information to dealers about the current state of the macroeconomy which they then use to revise their spot exchange rate quotes. Thus, the high frequency behavior of spot exchange rates reflects the flow of new information reaching dealers concerning the slowly evolving state of the macroeconomy, rather than the effects of shocks that drive rapidly changing macroeconomic conditions.

The third key element is empirical. To test the predictions of the model I need to quantify how dealers' views of macroeconomic conditions evolve. This is impossible using standard time series observations on macroeconomic variables. Instead I use a novel measure; the real-time error associated with estimating the current value of a macro variable (e.g. GDP) using a subset of information that was available to dealers at the time. These measures allow me to empirically test whether order flows actually carry information to dealers about the macro variable in question. The results from these tests are then used to quantify the degree to which excess returns embed the macro information conveyed to dealers via order flow.

The research I report on here builds on ideas developed in a series of papers by Richard Lyons and myself. In Evans and Lyons (2004) we first showed how the gap between traditional macro and microstructure exchange-rate models could be spanned by embedding the microfoundations of currency trading in a standard two-country DSGE model. This model demonstrated that equilibrium order flows convey macro information to dealers, but the information is insufficient to reveal the true state of the macroeconomy. In Evans and Lyons (2007) we developed the model further to study its implications for the forecasting power of order flows. Specifically, we identified the conditions under which order flows should have incremental forecasting power for the future path of macro variables that constitute exchange rate fundamentals; i.e., forecasting power beyond that provided by spot rates and other public information. This theoretical implication represents an extension of work by Engel and West (2005) based on traditional exchange-rate models, and receives strong empirical support in the data. The theoretical model I present here is a version of the model first developed in Evans and Lyons (2008b). There we derived and studied the equilibrium behavior of order flow, spot rates and the foreign exchange risk premium. Here I estimate the relation between order flows and macroeconomic variables in the data. Evans and Lyons (2008b) also differs from the current paper because its focus is on the origins of order flow's significant forecasting power for future currency returns.

My empirical results also relate to the findings of several other researchers, notably the work of Froot and Ramadorai (2005). They examine real exchange rates, excess currency returns, real interest differentials, and the transaction flows of institutional investors across 19 country/currency areas over seven years. Using variance decompositions from a VAR estimated at the daily frequency with pooled data, they find that innovations in transaction flows predict future changes in real interest differentials at short but not long

horizons. In contrast, the order flow data I use spans the full spectrum of agent-types not just institutional investors, and comes from USD/EUR transactions – the most liquid portion of the market. These facets of the flow data prove to be empirically important. My focus also differs from Froot and Ramadorai insofar as it concerns the incremental information order flow conveys to dealers concerning current macroeconomic conditions (e.g. output and inflation) rather than their potential forecasting power for future interest rates. My results are also related to recent research on the relation between order flows and macro data releases. Evans and Lyons (2005), Berger et al. (2005), Dominguez and Panthaki (2006) and Love and Payne (2007) all find that macro data releases have a statistically significant impact on order flows, while Evans and Lyons (2008a) estimate that one-third of the variance in order flow is attributable to the arrival of macroeconomic news. These results point to the presence of differing views among market participants concerning the current state of the economy and its implications for currency prices.[2] My analysis exploits these very features.

Empirical results always come with caveats. However, in light of the historically poor empirical performance of exchange-rate models and the novelty of my results, it is important to be up front about the caveats that apply here. My empirical findings are based on estimates of the flow of information concerning output, inflation and the money stocks from January 1993 to June 1999, conveyed by the world-wide flow of currency orders received by Citibank from end-users, such as corporations, mutual funds and hedge funds. These facets of my data give rise to two potential concerns. First, while Citibank was the single largest bank in the USD/EUR market at the time, its end-user flows represent between 10 and 15 percent of the flows received by banks across the entire market. This means that the information conveyed by Citibank's flows may not be entirely representative of the market-wide flow of information that drives spot rates. If this is the case, my estimates will represent a lower bound on the contribution of macro information flows to the variance of excess returns. The second concern relates to the data sample. Clearly, it is possible that order flows carried much more macro information during this period than at other times. However, in view of the well-documented empirical robustness of the relationship between spot rates and order flows over longer time spans and different currencies, there is no premia facie evidence to indicate significant instability in the underlying mechanism concerning order flow's role as a conveyor of information. Ultimately, of course, the robustness of my results will only be established by further research using longer and broader data samples. Since the collection of order flow data is now an on-going process, I hope that my results will serve as a stimulus for this research.

The remainder of the paper comprises 3 sections. Section 1 presents the theoretical model. Second 2 develops the implications of the model into set of empirically estimable relationships between exchange rates, order flows and macro data. I then describe the data and present the empirical results. Section 3 concludes.

---

[2] In contrast, if there is unanimous agreement on the implications of a macro data release for the value of foreign currency, rational dealers will adjust their quotes to fully reflect this view. Consequently, there should not be any systematic imbalance between orders to buy or sell foreign currency at this new spot rate. In short, the common knowledge information in the data release is immediately reflected in the spot rate but has no impact on order flow. For further discussion, see Evans and Lyons (2008a).

1 The Model

This section presents a micro-based model of exchange rate dynamics that identifies the role order flow plays in conveying macro information to the foreign exchange market. The model is micro-based in the sense that it incorporates key features of currency trading models, such as Evans and Lyons (1999 & 2004). In particular, the spot exchange rate is determined as the foreign currency price quoted by dealers who have limited information about the current state of the economy. The model also contains elements found in recent macro exchange-rate models (e.g., Mark 2005, and Engel and West 2006) insofar as dealers recognize that short-term interest rates are the policy instruments of central banks. Specifically, dealers' interest rate expectations incorporate a view on how central banks react to changes in the macroeconomy. The model's focus is on how dealers use order flow to draw inferences about the current state of macroeconomy, which in turn affect their interest-rate forecasts and their foreign currency quotes.

1.1 Structure

The economy comprises two countries populated by a continuum of risk-averse agents indexed by $n \in [0, 1]$, and a finite number of risk-averse dealers who act as market-makers in the spot market for foreign currency. I shall refer to the home and foreign countries as the US and Europe, so the log spot exchange rate, s_t, denotes the dollar price of euros. The only other actors in the model are the central banks (i.e. The Federal Reserve (FED) and the European Central Bank (ECB)), who conduct monetary policy by setting short-term nominal interest rates.

1.1.1 Dealers

The pattern of actual trading in the foreign exchange market is extremely complex. On the one hand, foreign exchange dealers quotes prices at which they stand ready to buy or sell foreign currency to agents and other dealers. On the other, each dealer can initiate trades against other dealers' quotes, and can submit both market and limit orders to electronic brokerages. Fortunately, for the task at hand, it is unnecessary to represent this trading activity in any detail. Instead, it suffices to focus on the price dealers quote at the start of each trading week. In particular, I assume that log spot price quoted by all dealers at the start of week t is given by

$$s_t = \mathbb{E}_t^{\mathrm{D}} s_{t+1} + \hat{r}_t - r_t - \delta_t, \tag{1}$$

where $\mathbb{E}_t^{\mathrm{D}}$ denotes expectations conditioned on the common information available to all dealers at the start of week t, Ω_t^{D}. This information set includes r_t and $\hat{r}_t$, which are the one-week dollar and euro interest rates set by the FED and ECB, respectively. (Hereafter I use hats, "ˆ", to denote European variables.) The last term on the right, δ_t, is a risk premium that dealers choose to manage risk efficiently. This risk premium is determined below as a function of dealers' common information, Ω_t^{D}.

In the currency trading models of Lyons (1997) and Evans and Lyons (1999 & 2004), the spot exchange rate is determined by the Perfect Bayesian Equilibrium (PBE) quote strategy of a game between the dealers played over multiple trading rounds. Equation (1) incorporates three features of the PBE quotes in these trading models: First, each dealer quotes the same price to agents and other dealers. Second, quotes are common across all dealers. Third, all quotes are a function of common information, Ω_t^{D}. Notice that

equation (1) does not implicitly restrict all dealers to have the same information. On the contrary, dealers will generally possess heterogenous information, which they use in forming their optimal trading strategies. However, insofar as our focus is on the behavior of the spot rate (rather than dealer trading), equation (1) implies that we can concentrate on dealers' common information, Ω_t^D.

Equation (1) says that the log spot price quoted by all dealers at the start of week t is equal to the expected payoff from holding foreign currency until the next week, $\mathbb{E}_t^D s_{t+1} + \hat{r}_t - r_t$, less a premium, δ_t. In models of currency trading, the size of this premium is determined by the requirements of efficient risk-sharing. More specifically, when there is a finite number of risk-averse dealers and a continuum of risk-averse agents, dealers will choose δ_t to minimize their expected holdings of risky currencies at the end of week t. To incorporate this risk-sharing implication, I assume that all dealers are located in the US. They therefore choose δ_t such that their expected holdings of euros at the end of week t equal zero. These holdings are determined by the history of order flows from all agents. In particular, let x_{t+1} denote the aggregate of all orders from agents for euros received by dealers during week t,[3] so $I_{t+1} = -\sum_{i=0}^{\infty} x_{t+1-i}$ denotes the euro holdings of all dealers at the end of week-t trading. Efficient risk-sharing requires that dealers chose a value for δ_t such that $\mathbb{E}_t^D I_{t+1} = 0$.[4] Clearly, this restriction makes δ_t a function of dealers' common information, Ω_t^D. It also accords well with the actual behavior of dealers, who are restricted on the size of their overnight positions (see, for example, Lyons 1995).

Recent exchange-rate models by Engel and West (2005 & 2006) and Mark (2005) have focused on the fact that central banks use short-term interest rates rather than monetary aggregates as policy instruments. In line with this research, I assume that dealers' interest rate expectations incorporate a view on how central banks react to changes in the macroeconomy. In particular, I assume that

$$\mathbb{E}_t^D(\hat{r}_{t+i} - r_{t+i}) = (1 + \gamma_\pi)\mathbb{E}_t^D(\Delta\hat{p}_{t+1+i} - \Delta p_{t+1+i}) + \gamma_y\mathbb{E}_t^D(\hat{y}_{t+i} - y_{t+i}) - \gamma_\varepsilon\mathbb{E}_t^D\varepsilon_{t+i}, \tag{2}$$

for $i > 0$, where γ_π, γ_y, and γ_ε are positive coefficients. Equation (2) says that dealers expect the future differential between euro and dollar rates to be higher when; (i) the future difference between EU and US inflation, $\Delta\hat{p}_{t+1} - \Delta p_{t+1}$, is higher, (ii) the difference between the EU and US output gaps, $\hat{y}_t - y_t$, widens, or (iii) when the real exchange rate, $\varepsilon_t \equiv s_t + \hat{p}_t - p_t$, depreciates. The first two terms are consistent with the widely-accepted view that central banks react to higher domestic inflation and output by raising short-term interest rates. The third term captures the idea that some central banks can be expected to react to deviations in the spot rate from its purchasing power parity level (i.e., the real exchange rate, ε_t); a notion that finds empirical support in Clarida, Gal, and Gertler (1998).

Equation (2) embodies an assumption about how dealers' expectations for future interest rates are related to their expectations concerning macro variables (e.g., inflation and output), rather than an assumption about whether central banks actually follow particular reaction functions, such as a Taylor-rule. As we shall see, the actions of central banks directly affect the spot rate via the current interest rates they set, but it is

---

[3] I identify the order flow from week-t trading with a subscript of $t + 1$ to emphasize the fact that dealers cannot use the information it conveys until the start of week $t + 1$.

[4] This implication of efficient risk-sharing also applies if dealers are distributed in both countries. In this case, I_{t+1} represents the US dealers' holding of euros minus the euro value of EU dealers' dollar holdings at the end of week-t trading. Efficient risk-sharing now requires that the expected value of the foreign currency holdings of all dealers are equalized, i.e. $\mathbb{E}_t^D I_{t+1} = 0$, after dealers have had the opportunity to trade with each other. By assuming that all dealers are located in the US, I am simply abstracting from the need to model intradealer trade.

dealer's expectations about how future interest rates are related to the macroeconomy that are central to understanding the role of order flow as a source of exchange rate dynamics.

Dealers have access to both private and public sources of information. Each dealer receives private information in the form of the currency orders from the subset of agents that trade with them, and from the currency orders they receive from other dealers. In currency trading models, the mapping from dealers' individual information sets to the common information set for all dealers, Ω_t^{D}, is derived endogenously from the trading behavior of dealers. I will not consider this complex process here. Instead, I characterize the evolution of Ω_t^{D} directly under the assumption that a week's worth of trading is sufficient to reveal the size of the aggregate order flow from agents to all dealers. Thus, all dealers know the aggregate order flow from week-t trading, x_{t+1}, by the start of week $t + 1$.

Dealers receive public information in the form of macro data releases and their observations on short-term interest rates. To characterize this information flow, let z_t denote a vector of variables that completely describes the state of the macroeconomy in week t. This vector contains short-term interest rates, r_t and $\hat{r}_t$; prices, p_t and $\hat{p}_t$; the output gaps, y_t and $\hat{y}_t$; and other variables. A subset of these variables, z_t^o, are contemporaneously observable to all dealers and agents. I assume that other elements of z_t only become publicly known via macro data releases with a reporting lag. The presence of the reporting lag is an important feature of the model and accords with reality. For example, data on US GDP in the first quarter is only released by The Bureau of Economic Analysis several weeks into the second quarter, so the reporting lag for US output can run to more than 16 weeks. Let $z_t^m = [z_{i,t}^m]$ denote the vector of macro data releases at the start of week t, where $z_{i,t}^m$ denotes the data release on variable i, (e.g., GDP). Each release represents an observation of *past* macroeconomic activity, so $z_{i,t}^m = z_{i,t-k_i}$ where k_i denotes the reporting lag for variable i. With these assumptions, the evolution of dealers' common information is given by

$$\Omega_t^{\text{D}} = \left\{ z_t^o, z_t^m, x_t, \Omega_{t-1}^{\text{D}} \right\}, \tag{3}$$

where $\{z_t^o, z_t^m\}$ identifies the source of the public information flow, and x_t identifies the source of the information flow observed by all dealers.

1.1.2 Agents and the Macroeconomy

Since my aim is to examine how the evolution of the macroeconomy is linked to activity in the foreign exchange market, there is no need to describe every aspect of agents' behavior. Instead, I focus on their demand for foreign currency. Specifically, I assume that the demand for euros in week t by agent $n \in [0, 1]$ is

$$\alpha_t^n = \alpha_s \left(\mathbb{E}_t^n \Delta s_{t+1} + \hat{r}_t - r_t \right) + h_t^n, \tag{4}$$

where $\alpha_s > 0$ and $\mathbb{E}_t^n$ denotes expectations conditioned on the information available to agent n after observing the spot rate at the start of week t, Ω_t^n. Equation (4) decomposes the demand for euros into two terms. The first is the (log) excess return expected by the agent, $\mathbb{E}_t^n \Delta s_{t+1} + \hat{r}_t - r_t$, the second is a hedging term, h_t^n, that represents the influence of all other factors. This representation of foreign currency demand is very general. For example, it could be derived from a mean-variance portfolio choice model, or from an OLG portfolio model such as in Bacchetta and van Wincoop (2006). In these cases, the h_t^n term identifies the

expected returns on other assets and the hedging demand induced by the exposure of the agent's future income to exchange-rate risk. Alternatively, the representation in (4) could be derived as an approximation to the optimal currency demand implied by an intertemporal portfolio choice problem, as in Evans and Hnatkovska (2007). In this case the h_t^n term would also incorporate the effects of variations in the agent's wealth. Without loss of generality, I assume that

$$h_t^n = \alpha_z z_t^n, \tag{5}$$

for some vector α_z, where z_t^n is a vector of variables that describes the microeconomic environment of agent n. This environment includes observable macro variables, such as interest rates, and the micro data that influences all aspects of the agent's behavior. Equation (5) simply states that the agent's hedging demand for foreign currency depends on his microeconomic environment. The link between this environment and the state of the macroeconomy is given by

$$z_t^n = z_t + v_t^n, \tag{6}$$

where $v_t^n = [v_{i,t}^n]$ is a vector of agent-specific shocks with the property that $\int_0^1 v_{i,t}^n dn = 0$ for all i.

Like dealers, each agent has access to both private and public sources of information. The former comes from his or her microeconomic environment, z_t^n. Each agent also receives public information about the macroeconomy from macro data releases, the short-term interest rates set by central banks, and from the spot exchange rate quoted by dealers. The evolution of agent n's information can therefore be represented as

$$\Omega_t^n = \left\{ z_t^o, z_t^m, s_t, z_t^n, \Omega_{t-1}^n \right\}, \tag{7}$$

for $n \in [0, 1]$.

All that now remains is to characterize the behavior of the macroeconomy. For this purpose, it suffices to identify a few elements of the z_t vector, and to represent its dynamics in a reduced form. Specifically, I assume that the interest, inflation, price and output differentials comprise the first four elements of z_t,

$$z_t' = [\ \hat{r}_t - r_t, \quad \Delta\hat{p}_t - \Delta p_t, \quad \hat{p}_t - p_t, \quad \hat{y}_t - y_t, \quad ..., \quad ... \ ],$$

and that the dynamics of z_t can be written as

$$z_t = A z_{t-1} + B u_t, \tag{8}$$

for some matrices A and B, where u_t is a vector of mean zero serially uncorrelated shocks. Equation (8) conveniently summarizes the aggregate implications of the optimal decisions of agents with respect to consumption, saving, investment, and price-setting in a manner consistent with market clearing given assumptions about productivity, preference shocks, and the conduct of monetary/fiscal policy.

1.2 The Equilibrium Spot Rate

In equilibrium, the spot rate quoted by dealers satisfies equation (1) subject to the risk-sharing restriction that identifies the risk premium, and dealers' expectations concerning future interest rates in (2). To derive

an expression for the equilibrium spot rate, I first use (1) to write

$$\mathbb{E}_t^{\mathrm{D}} s_{t+1} = \mathbb{E}_t^{\mathrm{D}} \left[\hat{r}_{t+1} - r_{t+1} - \delta_{t+1} \right] + \mathbb{E}_t^{\mathrm{D}} s_{t+2}.$$

Substituting for interest rate expectations with (2), and iterating forward assuming that $\mathbb{E}_t^{\mathrm{D}} \lim_{i \to \infty} \rho^i s_{t+i} = 0$ gives

$$\mathbb{E}_t^{\mathrm{D}} s_{t+1} = \mathbb{E}_t^{\mathrm{D}} \sum_{i=1}^{\infty} \rho^i (f_{t+i} - \delta_{t+i}), \tag{9}$$

with $\rho \equiv 1/(1 + \gamma_\varepsilon) < 1$, where $f_t = \left[\gamma_\pi \left(\Delta \hat{p}_{t+1} - \Delta p_{t+1} \right) + \gamma_y \left(\hat{y}_t - y_t \right) \right] + \left(\frac{1-\rho}{\rho} \right) (p_t - \hat{p}_t)$. Equation (9) identifies dealers' expectations for next week's spot rate in terms of their forecasts for macro fundamentals, f_t, and the risk premium, δ_t. Combining (9) with (1) gives the following equation for the equilibrium spot rate:

$$s_t = (\hat{r}_t - r_t) + \mathbb{E}_t^{\mathrm{D}} \sum_{i=1}^{\infty} \rho^i f_{t+i} - \mathbb{E}_t^{\mathrm{D}} \sum_{i=0}^{\infty} \rho^i \delta_{t+i}, \tag{10}$$

The three terms on the right of equation (10) identify different factors affecting the log spot rate dealers quote at the start of week t. First, the current stance of monetary policy in the US and EU affects dealers' quotes via the interest differential, $\hat{r}_t - r_t$, because it directly contributes to the payoff from holding euros until week $t+1$. Second, dealers are concerned with the future course of macro fundamentals, f_t. This term embodies dealers' expectations of how central banks will react to macroeconomic conditions when setting future interest rates. The third factor arises from risk-sharing between dealers and agents as represented by the present and expected future values of the risk premium. This risk-sharing implication is not found in standard exchange-rate models, and so requires some further explanation.

Recall that dealers choose the risk premium so that $\mathbb{E}_t^{\mathrm{D}} I_{t+1} = 0$ where $I_{t+1} = I_t - x_{t+1}$. Now market clearing requires that the aggregate order flow received by dealers during week-t trading must equal the aggregate change in the demand for euros across all agents: $x_{t+1} = \alpha_t - \alpha_{t-1}$ where $\alpha_t = \int_0^1 \alpha_t^n dn$ is the aggregate demand for euros. These two conditions jointly imply that $I_{t+1} + \alpha_t = I_t + \alpha_{t-1} = I_1 + \alpha_0$. For clarity, I normalize $I_1 + \alpha_0$ to equal zero, so efficient risk-sharing requires that $0 = \mathbb{E}_t^{\mathrm{D}} \alpha_t$. Combining this restriction with (4) - (6) and the definition of α_t gives

$$0 = \alpha_s \mathbb{E}_t^{\mathrm{D}} (\overline{\mathbb{E}}_t^n s_{t+1} - s_t + \hat{r}_t - r_t) + \mathbb{E}_t^{\mathrm{D}} h_t, \tag{11}$$

where $h_t = \int_0^1 h_t^n dn$ is the aggregate hedging demand for euros and $\overline{\mathbb{E}}_t^n s_{t+1} = \int_0^1 \mathbb{E}_t^n s_{t+1} dn$ is the average of agents' forecasts for next week's spot rate. Since $\mathbb{E}_t^{\mathrm{D}} \Delta s_{t+1} + \hat{r}_t - r_t = \delta_t$ from equation (1), and $\mathbb{E}_t^{\mathrm{D}} \overline{\mathbb{E}}_t^n s_{t+1} = \mathbb{E}_t^{\mathrm{D}} s_{t+1} + \mathbb{E}_t^{\mathrm{D}} (\overline{\mathbb{E}}_t^n s_{t+1} - s_{t+1})$ by definition, the first term on the right of (11) is equal to $\alpha_s \delta_t + \alpha_s \mathbb{E}_t^{\mathrm{D}} (\overline{\mathbb{E}}_t^n s_{t+1} - s_{t+1})$. We may therefore rewrite equation (11) as

$$\delta_t = \mathbb{E}_t^{\mathrm{D}} [s_{t+1}^e - \tfrac{1}{\alpha_s} h_t], \tag{12}$$

where $s_{t+1}^e \equiv s_{t+1} - \overline{\mathbb{E}}_t^n s_{t+1}$.

Equation (12) shows that dealers' choice for the risk premium depends on their estimates of: (i) the aggregate hedging demand for euros, $\mathbb{E}_t^{\mathrm{D}} h_t$; and (ii) the average error agents make when forecasting next

week's spot rate, s_{t+1}^e. Intuitively, dealers lower the risk premium when they anticipate a rise in the aggregate hedging demand for euros because the implied fall in the excess return agents expect will offset their desire to accumulate larger euro holdings. Dealers also reduce the risk premium to offset agents' desire to accumulate larger euro holdings when they are viewed as being too optimistic (on average) about the future spot rate; that is when $\mathbb{E}_t^{\mathrm{D}} s_{t+1} < \mathbb{E}_t^{\mathrm{D}} \overline{\mathbb{E}}_t^n s_{t+1}$.

Combining (10) with (12) gives us the following micro-based exchange rate equation:

$$s_t = (\hat{r}_t - r_t) + \mathbb{E}_t^{\mathrm{D}} \sum_{i=1}^{\infty} \rho^i f_{t+i} + \tfrac{1}{\alpha_s} \mathbb{E}_t^{\mathrm{D}} \sum_{i=0}^{\infty} \rho^i h_{t+i} - \tfrac{1}{\rho} \mathbb{E}_t^{\mathrm{D}} \sum_{i=1}^{\infty} \rho^i s_{t+i}^e. \tag{13}$$

Two aspects of equation (13) differentiate the exchange-rate implications of this model from those of traditional macro models. First, the equation identifies the foreign currency price all dealers quote. As such, s_t is a function of the information available to dealers, Ω_t^{D}, at the time they quote prices. This information set includes contemporaneous interest rates, $\hat{r}_t$ and r_t, but not observations on the other variables that comprise the current state of the macroeconomy. Second, dealers' quotes include a risk premium that incorporates their estimates of aggregate hedging demand and agents' forecast errors. These factors can be a source of variation in spot rates when dealers' views of current and future monetary policy remain unchanged.[5]

1.3 Equilibrium Dynamics

The dynamic implications of the model differ from those found in macro exchange-rate models. I now develop these implications to explain the role order flow plays in conveying macro information to the foreign exchange market.

Consider the implications of (13) for the behavior of log excess returns: $er_{t+1} \equiv \Delta s_{t+1} + \hat{r}_t - r_t$. Equation (1) implies that $er_{t+1} = \delta_t + s_{t+1} - \mathbb{E}_t^{\mathrm{D}} s_{t+1}$, so substituting for the risk premium with (12) and dealers' forecast error with (13) gives

$$er_{t+1} = \mathbb{E}_t^{\mathrm{D}} [s_{t+1}^e - \tfrac{1}{\alpha_s} h_t] + (1 - \mathbb{E}_t^{\mathrm{D}})(\hat{r}_{t+1} - r_{t+1}) + \eta_{t+1} - \omega_{t+1}, \tag{14}$$

where

$$\eta_{t+1} = (\mathbb{E}_{t+1}^{\mathrm{D}} - \mathbb{E}_t^{\mathrm{D}}) \sum_{i=2}^{\infty} \rho^{i-1}(f_{t+i} + \tfrac{1}{\rho \alpha_s} h_{t+i-1}) \qquad \text{and} \qquad \omega_{t+1} = (\mathbb{E}_{t+1}^{\mathrm{D}} - \mathbb{E}_t^{\mathrm{D}}) \sum_{i=2}^{\infty} \rho^{i-2} s_{t+i}^e.$$

Equation (14) relates excess returns to dealers' information. The first term on the right identifies the risk premium, δ_t, as a function of dealers' common information at the start of week t, Ω_t^{D}. The remaining terms depend on the flow of information reaching all dealers between the start of weeks t and $t+1$. This information flow comprises observations on current macro variables, z_{t+1}^o (e.g. $\hat{r}_{t+1} - r_{t+1}$), data releases on *past* macro variables, z_{t+1}^m, and order flow from week$-t$ trading, x_{t+1}. Equation (14) says that they affect the excess

---

[5] The presence of the last term in equation (13) is reminiscent of the rational expectations exchange rate models developed by Bacchetta and van Wincoop (2006 & 2008). In those models, the market clearing spot rate depends on current fundamentals and higher-order expectations concerning future macro fundamentals. Bacchetta and van Wincoop (2008) show that the latter can be written in terms of the average expectation of future average forecast errors. Here it is dealers' expectations of the agents' average forecast errors that affect the spot rate via their implications for efficient risk-sharing.

returns insofar as they convey new information on current monetary policy, $\hat{r}_{t+1} - r_{t+1}$; future fundamentals and hedging demand, η_{t+1}; and the future forecasts errors of agents, ω_{t+1}.

To study the relation between the sources and uses of dealers' information, it proves useful to rewrite (14) in terms of the vector of macro variables, z_t, and their dynamics in (8). Specifically, let the vectors λ_r and λ_f select the elements in $\hat{r}_t - r_t$ and f_t from z_t, (i.e., $\hat{r}_t - r_t = \lambda_r z_t$ and $f_t = \lambda_f z_t$). Equation (14) can now be rewritten as

$$er_{t+1} = \delta_t - \omega_{t+1} + \lambda(\mathbb{E}_{t+1}^{\mathrm{D}} - \mathbb{E}_t^{\mathrm{D}})z_{t+1}, \tag{15}$$

where $\lambda = \lambda_r + (\lambda_f A + \frac{1}{\rho\alpha_s}\alpha_z)\rho(I - \rho A)^{-1}$. Dealers use the flow of information between the start of weeks t and $t+1$ to revise their estimates about: (i) the future forecast errors of agents, ω_{t+1}; and (ii) the state of the macroeconomy in week $t+1$, $(\mathbb{E}_{t+1}^{\mathrm{D}} - \mathbb{E}_t^{\mathrm{D}})z_{t+1}$. In the former case, dealers only use the information contained in order flow. The reason is that the average forecast error, $s_{t+i}^e \equiv s_{t+i} - \overline{\mathbb{E}}_{t+i-1}^n s_{t+i}$ for $i \geq 1$, cannot be correlated with week$-t$ public information (i.e., elements of $\Omega_t = \cap_n \Omega_t^n$). This means that dealers' expectations, $\mathbb{E}_t^{\mathrm{D}} s_{t+i}^e$, can only differ from zero insofar as Ω_t^{D} contains relevant non-public information. As a consequence, any revisions in dealers expectations, $(\mathbb{E}_{t+1}^{\mathrm{D}} - \mathbb{E}_t^{\mathrm{D}})s_{t+i}^e$, must be a function of unexpected order flow from week$-t$ trading, $x_{t+1} - \mathbb{E}_t^{\mathrm{D}} x_{t+1}$, because this is the only source of non-public information flow available to all dealers.

The flows of both public and non-public information can contribute to the last term in (15); the revision in dealers estimates concerning the macroeconomy, $(\mathbb{E}_{t+1}^{\mathrm{D}} - \mathbb{E}_t^{\mathrm{D}})z_{t+1}$. To see why, let $z_{t+1}^{\mathrm{D}} = [(z_{t+1}^o)', (z_{t+1}^m)', x_{t+1}]'$ denote the vector of observations all dealers receive between the start of weeks t and $t+1$. Furthermore, and without loss of generality, let the z_{t+1} vector contain the lagged values of the macro variables that are reported in the data releases together with all the variables that determine agents' aggregate demand for euros in weeks t and $t-1$, α_t and α_{t-1}. Using the market clearing condition, $x_{t+1} = \alpha_t - \alpha_{t-1}$, we can then write $z_{t+1}^{\mathrm{D}} = C z_{t+1}$ for some matrix C.[6] Applying the Kalman Filter to this expression and the dynamics of z_{t+1} in (8) gives

$$(\mathbb{E}_{t+1}^{\mathrm{D}} - \mathbb{E}_t^{\mathrm{D}})z_{t+1} = G(z_{t+1}^{\mathrm{D}} - \mathbb{E}_t^{\mathrm{D}} z_{t+1}^{\mathrm{D}}) = GCB u_{t+1} + GCA(z_t - \mathbb{E}_t^{\mathrm{D}} z_t), \tag{16}$$

where $G = \mathbb{V}_t^{\mathrm{D}}(z_{t+1})C'(C\mathbb{V}_t^{\mathrm{D}}(z_{t+1})C')^{-1}$ is the filter's gain matrix and $\mathbb{V}_t^{\mathrm{D}}(z_{t+1})$ is the covariance of z_{t+1} conditioned on Ω_t^{D}.

Equation (16) shows that dealers revise their estimates of z_{t+1} in response to information about current shocks, u_{t+1}, and the *past* state of the macroeconomy, $z_t - \mathbb{E}_t^{\mathrm{D}} z_t$. The former is conveyed by dealers' observations on contemporaneous variables, such as the interest rates set by central banks. Equation (15) identifies the effect of such shocks on the spot rate as $\lambda GCB u_{t+1}$. It is through this channel that unexpected macro developments affect spot rates in macro exchange-rate models.

Macro data releases, z_{t+1}^m, and order flow, x_{t+1}, contribute to $(\mathbb{E}_{t+1}^{\mathrm{D}} - \mathbb{E}_t^{\mathrm{D}})z_{t+1}$ insofar as they convey

---

[6]It is important to emphasize that relating dealers' observations to the z_{t+1} vector in this way does *not* embody any assumption concerning the dependence of the aggregate demand for euros on specific macro variables. In particular, even if the equilibrium demand for euros depends on a sub-vector of variables in z_{t+1} that are uncorrelated with the variables that comprise fundamentals, f_t, it is still possible to write α_t and α_{t-1} in terms of z_{t+1}. Evans and Lyons (2008b) derive the equilibrium expression for α_t in terms of z_{t+1} under the assumption that the vector z_t becomes known to all dealers and agents with a finite reporting lag (i.e., $z_{t-k} \in \Omega_t$ and $z_{t-k} \in \Omega_t^{\mathrm{D}}$ for some $k < \infty$). This paper adopts the complementary strategy of empirically examining the link between agents' aggregate demands and key macro variables.

new information about z_t. Each data release represents an observation of *past* macroeconomic activity, so $z_{t+1}^m - \mathbb{E}_t^{\mathrm{D}} z_{t+1}^m$ simply picks out an element of $z_t - \mathbb{E}_t^{\mathrm{D}} z_t$. (Recall that z_t contains the values of macro variables in week t and earlier.) Order flow also conveys information about $z_t - \mathbb{E}_t^{\mathrm{D}} z_t$. Since dealers know the history of order flow and $\alpha_{t-1} = \sum_{i=0}^{\infty} x_{t-i}$ by market clearing, $\alpha_{t-1} \in \Omega_t^{\mathrm{D}}$. Consequently, unexpected order flow from week-t trading is $x_{t+1} - \mathbb{E}_t^{\mathrm{D}} x_{t+1} = (\alpha_t - \alpha_{t-1}) - \mathbb{E}_t^{\mathrm{D}} (\alpha_t - \alpha_{t-1}) = \alpha_t - \mathbb{E}_t^{\mathrm{D}} \alpha_t$. Substituting for α_t in this expression using (4), (5), (6) and (12) gives

$$x_{t+1} - \mathbb{E}_t^{\mathrm{D}} x_{t+1} = \alpha_z (z_t - \mathbb{E}_t^{\mathrm{D}} z_t) + \alpha_s (\overline{\mathbb{E}}_t^n s_{t+1} - \mathbb{E}_t^{\mathrm{D}} [\overline{\mathbb{E}}_t^n s_{t+1}]). \tag{17}$$

Thus order flow contributes new information about z_t, via the aggregate hedging demand, $h_t - \mathbb{E}_t^{\mathrm{D}} h_t = \alpha_z (z_t - \mathbb{E}_t^{\mathrm{D}} z_t)$, and about the average of agents' spot rate forecasts, $\overline{\mathbb{E}}_t^n s_{t+1}$.

Equation (17) provides the key to understanding why order flow has the potential to convey timely macro information to dealers. In particular, notice that unexpected order flow during week t contains information on agent's aggregate hedging demands, $h_t = \alpha_z \int_0^1 z_t^n dt$, and their spot rate forecasts, $\overline{\mathbb{E}}_t^n s_{t+1}$. Both of these factors depend on the microeconomic environments agents face in week t. As a consequence, order flow during week t carries more timely information about the current state of the economy than is available from the most recent macro data releases.[7]

We can now provide a structural interpretation of the relationship between excess returns and order flow. Substituting (16) into (15) gives the following expression for excess returns:

$$er_{t+1} = \delta_t + \lambda_u u_{t+1} + \lambda_z (z_t - \mathbb{E}_t^{\mathrm{D}} z_t) - \omega_{t+1}, \tag{18}$$

where $\lambda_u = \lambda GCB$ and $\lambda_z = \lambda GCA$. The first two terms on the right are uncorrelated with order flow, $x_{t+1} - \mathbb{E}_t^{\mathrm{D}} x_{t+1}$, because the risk premium is a function of dealers' common information, Ω_t^{D}, and $\lambda_u u_{t+1}$ identifies the effects of public macro shocks that were unanticipated by dealers and agents. In contrast, order flow can be correlated with either the third or forth terms. Order flow will be correlated with the third term when it provides macro information useful to dealers in revising their forecasts about future monetary policy and hedging demand, η_{t+1}. It will be correlated with the forth term when dealers use the information in order flow to revise their estimates of agents' future forecasts errors. In this case order flow need not convey any information concerning conventional macroeconomic variables such as output and inflation. Agents' forecast errors could be solely a function of private information (i.e. elements in z_t^n) that are completely uncorrelated with conventional macro variables. Equation (17) shows that order flow aggregates this information, and so could be useful to dealers in revising their estimates of agents' future forecast errors, ω_{t+1}.

This micro-based exchange rate model provides a much richer picture of the link between the dynamics of spot rates and macroeconomic developments than is found in standard models. In particular, because the model identifies the spot rate as the price for foreign currency quoted by dealers, the behavior of excess returns reflects how dealers change their quote decisions week-by-week. This means that macroeconomic developments will only be reflected in the dynamics of spot rates insofar as they are observable to dealers

---

[7] All macro data realeases, except Civilian Unemployment, contain a reporting lag of more than one week; and even this exception does not provide complete information on z_t.

in real time, and contain price-relevant information. These are non-trivial requirements but they are largely ignored in macro exchange-rate models. For example, traditional monetary models and more recent Taylor rule models assume that all the macro variables determining interest differentials and the foreign exchange risk premium are contemporaneously and publicly observed. This assumption is not only counter-factual, but it also rules out some potentially important channels that link the dynamics of the spot rate to macroeconomic developments. Specifically, since $\mathbb{E}_t^D s_{t+1}^e = 0$ and $z_t = \mathbb{E}_t^D z_t$ under these circumstances, the last two terms in equation (18) disappear and $\delta_t = -\frac{1}{\alpha_s} h_t$. This means that all the variation in excess returns is driven by changes in aggregate hedging demand, h_t, and shocks to the macroeconomy, u_{t+1}. The exchange-rate disconnect puzzle (Obstfeld and Rogoff 2001) attests to the lack of empirical success in identifying either h_t or u_{t+1} with macro variables in the past 25 years of exchange-rate research. In contrast, the dealer-centered view developed here has the potential to reconnect spot rate dynamics to macroeconomic developments because order flow can convey price-relevant macro information to dealers who have limited information concerning the current state of the economy.

2   Empirical Analysis

This section empirically examines whether order flow conveys macro information about the current state of the economy that dealers then incorporate into their spot rate quotes. For this purpose, I first develop the implications of the model into a set of empirically estimable relationships between exchange rates, order flows and macro data. I then describe the data used to estimate these relationships and the estimation results.

2.1   Identifying the Information Conveyed by Order Flow

Consider the implications of the model for the projection of excess returns on unexpected order flow:

$$er_{t+1} = \beta_{er}(x_{t+1} - \mathbb{E}_t^D x_{t+1}) + \xi_{t+1}. \tag{19}$$

The projection coefficient is given by

$$\beta_{er} = \frac{\mathbb{E}[er_{t+1}(x_{t+1} - \mathbb{E}_t^D x_{t+1})]}{\mathbb{E}[(x_{t+1} - \mathbb{E}_t^D x_{t+1})^2]},$$

where $\mathbb{E}[.]$ denotes unconditional expectations. The portion of excess returns that is uncorrelated with unexpected order flow is identified by the projection error, ξ_{t+1}. Equation (18) implies that the projection coefficient can be decomposed as

$$\beta_{er} = \sum_{i=1}^{q} \lambda_{z_i} \beta_{z_i} - \beta_{\omega}, \tag{20}$$

where

$$\beta_{z_i} = \frac{\mathbb{E}[(z_{i,t} - \mathbb{E}_t^D z_{i,t})(x_{t+1} - \mathbb{E}_t^D x_{t+1})]}{\mathbb{E}[(x_{t+1} - \mathbb{E}_t^D x_{t+1})^2]}, \qquad \beta_{\omega} = \frac{\mathbb{E}[\omega_{t+1}(x_{t+1} - \mathbb{E}_t^D x_{t+1})]}{\mathbb{E}[(x_{t+1} - \mathbb{E}_t^D x_{t+1})^2]},$$

and λ_{z_i} denote the q elements of the λ_z vector.

We can use equation (20) to evaluate whether order flows convey macro information to dealers. Suppose, for example, that order flow only conveys information on agents' hedging demands that is uncorrelated with

the standard macro variables which make up the first k elements of z_t. Under these circumstances, $\beta_{z_i} = 0$ for $i \leq k$, so (20) becomes $\beta_{er} = \sum_{i=k+1}^{q} \lambda_{z_i} \beta_{z_i} - \beta_\omega$. Here excess returns are correlated with order flow, as we observe in the data, but this observation does not signify the presence of any connection between spot rates and standard macro variables. Alternatively, suppose that order flow aggregates dispersed information on the microeconomic environments agents face that is strongly correlated with the standard macro variables. Because this information has yet to be made public via macro data releases, β_{z_i} will differ from zero for $1 \leq i \leq k$. Under these circumstances, order flow's role as a conveyor of macro information will contribute to observed relation between excess returns and order flow via $\sum_{i=1}^{k} \lambda_{z_i} \beta_{z_i}$ provided the information is price-relevant in the sense that the λ_{z_i} coefficients differ from zero.

I empirically examine whether order flow conveys price-relevant macro information in two steps: First, I obtain estimates of β_{z_i} for a set of macro variables, z_i. These estimates quantify the extend to which order flow conveys new information to dealers concerning the current value of $z_{i,t}$. Second, I estimate the λ_{z_i} parameters from a regression of excess returns on the estimates of $\beta_{z_i}(x_{t+1} - \mathbb{E}_t^D x_{t+1})$ obtained from the first step. These estimates quantify whether dealers find the macro information conveyed by order flow useful in revising their spot rate quotes.

Before I implement these steps, two data issues need addressing. Notice that all the projection coefficients, β_{er}, β_ω and β_{z_i}, contain moments involving unanticipated order flow, $x_{t+1} - \mathbb{E}_t^D x_{t+1}$. It therefore appears that we need data on both x_{t+1} and dealers' information, Ω_t^D, in order to estimate β_{z_i} in the first step. Fortunately, another implication of the model makes this unnecessary. Recall that dealers choose the risk premium such that $\mathbb{E}_t^D \alpha_t = 0$, and $x_{t+1} - \mathbb{E}_t^D x_{t+1} = \alpha_t - \mathbb{E}_t^D \alpha_t$ because $\alpha_{t-1} \in \Omega_t^D$. Combining these expressions with the market clearing condition, $\alpha_t = \alpha_{t-1} + x_{t+1}$, gives $x_{t+1} - \mathbb{E}_t^D x_{t+1} = \sum_{i=0}^{\infty} x_{t+1-i}$. Thus, the requirement of efficient risk-sharing on the dealers choice of risk premium implies that unexpected order flow can be identified from the cumulation of current and past order flows, $\tilde{x}_{t+1}$.

The second issue concerns the identification of $z_{i,t} - \mathbb{E}_t^D z_{i,t}$. Again, we do not have data on dealers' information, Ω_t^D, so we cannot estimate β_{z_i} directly from the sample moments involving $z_{i,t} - \mathbb{E}_t^D z_{i,t}$. In principle we can circumvent this problem by noting that $\mathbb{E}[\tilde{x}_{t+1} \mathbb{E}_t^D z_{i,t}] = 0$, so $\beta_{z_i} = \mathbb{E}[\tilde{x}_{t+1} z_{i,t}]/\mathbb{E}[\tilde{x}_{t+1}^2]$. This suggests that β_{z_i} could be estimated from a regression of the macro variable, $z_{i,t}$, on the cumulation of order flow, $\tilde{x}_{t+1}$. Unfortunately, this approach is unlikely to have much statistical power in practice because the available order flow data covers six and a half years. This is a long time span from a trading perspective, but it does not contain many observations on standard macro variables such as output, inflation and monetary growth across a variety of macroeconomic conditions. Consequently, the available time series on $z_{i,t}$ are unlikely to be very informative about the information content of order flow.

To address this issue, I estimate β_{z_i} using real-time estimation errors for each macro variable constructed as $e_{i,t}^z = z_{i,t} - \mathbb{E}[z_{i,t}|\Omega_t]$, where Ω_t is a subset of dealers' common information, Ω_t^D. Combining the definition of β_{z_i} with the identity, $z_{i,t} - \mathbb{E}_t^D z_{i,t} \equiv e_{i,t}^z + \mathbb{E}[z_{i,t}|\Omega_t] - \mathbb{E}_t^D z_{i,t}$, gives

$$\beta_{z_i} = \frac{\mathbb{E}[e_{i,t}^z \tilde{x}_{t+1}]}{\mathbb{E}[\tilde{x}_{t+1}^2]} + \frac{\mathbb{E}[(\mathbb{E}[z_{i,t}|\Omega_t] - \mathbb{E}_t^D z_{i,t})\tilde{x}_{t+1}]}{\mathbb{E}[\tilde{x}_{t+1}^2]} = \frac{\mathbb{E}[e_{i,t}^z \tilde{x}_{t+1}]}{\mathbb{E}[\tilde{x}_{t+1}^2]},$$

because Ω_t is a subset of Ω_t^D. So estimates of β_{z_i} can be obtained from the regression of the real-time estimation errors on the cumulation of order flow. As we shall see, the real-time estimation errors display a significant degree of variability over the data sample. Consequently, their use allows me to examine the

information conveyed by order flow with much greater statistical precision than would be possible if we used standard macro time-series.

2.2 Data

My empirical analysis utilizes order flows, spot rates and real-time estimation errors for six macro variables over six and a half years; January 1993 to June 1999. The order flow data come from transactions between end-user (i.e. agents) and a large bank, Citibank. These data are disaggregated into six segments: trades executed between Citibank and non-financial corporations, investors (such as mutual funds and pension funds), and leveraged traders (such as hedge funds and proprietary traders) in the US; and trades executed between these same three groups of end-users and Citibank outside the US. Because it is possible for end-users in any location to trade with Citibank at any of its world-wide offices, these data do not provide information on the location of end-users. Nevertheless, it is reasonable to assume that the trade location is correlated with the location of the end-user – an inference that appears consistent with the estimation results reported below. The order flow data include all the end-user trades with Citibank in the largest spot market, the USD/EUR market, and the USD/EUR forward market.[8] Citibank had the largest share of the end-user market in these currencies at the time, ranging between 10 and 15 percent. The flow data measure in millions of dollars the imbalance between end-user orders to purchase and sell euros.

I examine the joint behavior of excess returns, order flows and the real-time estimation errors at a weekly frequency. The weekly timing of the variables is as follows: I take the log spot rate at the start of week t, s_t, to be the log of the offer rate (USD/EUR) quoted by Citibank at the end of trading on Friday of week $t-1$ (approximately 17:00 GMT). This is also the point at which I sample the week$-t$ interest rates from Datastream. Log excess returns over the τ-week horizon are computed as $er_{t+\tau}^\tau = s_{t+\tau} - s_t + \hat{r}_t^\tau - r_t^\tau$ where $\hat{r}_t^\tau$ and r_t^τ are the τ-week rates on euro-deutschmark and euro-dollar deposits. The week-t flow from segment j, x_j, is computed as the total dollar value of euro purchases minus sales initiated by the segment against Citibank's quotes between the 17:00 GMT on Friday of week $t-1$ and Friday of week t. Positive values for these order flows therefore denote net demand for euros by the end-user segment.

Summary statistics for the weekly order flow data are reported in Table 1. The statistics in panel A display two noteworthy features. First, the order flows are large and volatile. Second, they display no significant serial correlation. At the weekly frequency, the end-user flows appear to represent shocks to the foreign exchange market arriving at Citibank. Panel B reports the cross-correlations between the six flows. These correlations are generally quite small, ranging from approximately -0.16 to 0.16, but several are statistically significant at the 5 percent level. Insofar as these order flows convey information to dealers, individual segments should not be view as carrying entirely separate information.

I use the real-time estimation errors for GDP, the CPI, and M1 in the US and Germany to study the macro information conveyed by order flow. To understand how these errors are computed, let $z_{i,\text{M}(\mathfrak{m})}$ denote the value of a variable z_i representing macroeconomic activity during month $\mathfrak{m}$, that ends on day $\text{M}(\mathfrak{m})$. The value of z_i is released after the end of month $\mathfrak{m}$ and so contains a reporting lag. Reporting lags vary in length because data is collected on a calendar basis, but releases issued by statistical agencies are not

---

[8]Before January 1999, data for the Euro are synthesized from data in the underlying markets against the dollar, using weights of the underlying currencies in the euro.

Table 1: Order Flow Summary Statistics

| | Mean | Max | Skewness | Autocorrelations | | | |
	Std.	Min	Kurtosis	ρ_1	ρ_2	ρ_4	ρ_8
A:							
(i) Corporate US	-16.774	549.302	-0.696	-0.037	-0.040	0.028	-0.028
	108.685	-529.055	9.246	(0.434)	(0.608)	(0.569)	(0.562)
(ii) Corporate Non-US	-59.784	634.918	-0.005	0.072	0.089	-0.038	0.103
	196.089	-692.419	3.908	(0.223)	(0.124)	(0.513)	(0.091)
(iii) Traders US	-4.119	1710.163	0.026	-0.021	0.024	0.126	-0.009
	346.296	-2024.275	8.337	(0.735)	(0.602)	(0.101)	(0.897)
(iv) Traders Non-US	11.187	972.106	0.392	-0.098	0.024	0.015	0.083
	183.36	-629.139	5.86	(0.072)	(0.660)	(0.747)	(0.140)
(v) Investors US	19.442	535.320	-1.079	0.096	-0.024	-0.030	-0.016
	146.627	-874.15	11.226	(0.085)	(0.568)	(0.536)	(0.690)
(vi) Investors Non-US	15.85	1881.284	0.931	0.061	0.107	-0.030	-0.014
	273.406	-718.895	9.253	(0.182)	(0.041)	(0.550)	(0.825)

| B: Cross-Correlations | Corporate | | Hedge | | Investor | |
	US	Non-US	US	Non-US	US	Non-US
(i) Corporate US						
(ii) Corporate Non-US	-0.084					
(iii) Traders US	0.125*	-0.136*				
(iv) Traders Non-US	0.035	-0.026	0.066			
(v) Investors US	-0.158*	0.035	0.045	0.083		
(vi) Investors Non-US	-0.029	-0.063	0.159*	-0.032	0.094	

Notes: The table reports summary statistics for order flows from end-user segments cumulated over a week at the weekly frequency between January 1993 and June 1999. The last four columns on the right of panel A report autocorrelations ρ_i at lag i and p-values for the null that $\rho_i = 0$ in parentheses. Statistical significance at the 5% level is denoted by *.

made on holidays and weekends. (For quarterly series, such as GDP, reporting lags can be as long as several months.) The real-time estimation error for z_i on day $\eth$ in month $\mathfrak{m}$ is the difference between $z_{i,\mathrm{M}(\mathfrak{m})}$ and its estimated value based on a subset of dealers' common information on day $\eth$, $\Omega_\eth$. Formally, the real-time estimation error of a monthly series z_i is

$$z_{i,\mathrm{M}(\mathfrak{m})} - \mathbb{E}[z_{i,\mathrm{M}(\mathfrak{m})}|\Omega_\eth] \qquad \text{for } \mathrm{M}(\mathfrak{m}-1) < \eth \leq \mathrm{M}(\mathfrak{m}). \tag{21}$$

In the case of a quarterly series like GDP, the real-time estimation error on day $\eth$ is

$$z_{i,\mathrm{Q}(i)} - \mathbb{E}[z_{i,\mathrm{Q}(i)}|\Omega_\eth] \qquad \text{for } \mathrm{Q}(\mathfrak{q}-1) < \eth \leq \mathrm{Q}(\mathfrak{q}), \tag{22}$$

where $\mathrm{Q}(\mathfrak{q})$ denotes the last day of quarter $\mathfrak{q}$. I examine the macro information conveyed by order flow using the real-time estimation errors at the start of week t, computed from a subset of the information available to dealers at the end of Friday trading in week $t-1$. For the US variables, the information set, $\Omega_\eth$, includes the 3 quarterly releases on US GDP and the monthly releases on 18 other US macro variables. The real-

time estimation errors for German variables are computed using a specification for $\Omega_{\mathfrak{d}}$ that includes the 3 quarterly release on German GDP and the monthly releases on 8 German macro variables.[9] All series come from a database maintained by Money Market News Services (M.M.S.) that contains details of each data release. The real-time errors are computed with the method developed in Evans (2005).

Summary statistics for the real-time estimation errors, measured in percentage terms, are reported in Table 2. Two features stand out in Panel A. First, the real-time errors are highly variable, except in the case of the US CPI. This reflects the fact that the real-time estimates, $\mathbb{E}[z_{i,\mathrm{M(m)}}|\Omega_{\mathfrak{d}}]$ and $\mathbb{E}[z_{i,\mathrm{Q}(i)}|\Omega_{\mathfrak{d}}]$, vary significantly from week to week with the flow of macro data releases. The second feature concerns the autocorrelations. These are large, positive and highly statistically significant over horizons of one to eight weeks. This feature arises because the macro data releases used to construct the real-time estimates contain significant reporting lags. As a consequence, it takes several weeks before these releases reveal precise information about the macro variable under consideration.

To see this more clearly, consider the autocovariance for the real-time estimation error with a k-week lag: $\mathbb{CV}(e_{i,t}^{z}, e_{i,t-k}^{z})$. By definition, $e_{i,t}^{z} \equiv e_{i,t-k}^{z} + (\Delta^{k} z_{i,t} - \mathbb{E}[\Delta^{k} z_{i,t}|\Omega_t]) - (\mathbb{E}[z_{i,t-k}|\Omega_t] - \mathbb{E}[z_{i,t-k}|\Omega_{t-k}])$, so we can rewrite this autocovariance as

$$\mathbb{CV}(e_{i,t}^{z}, e_{i,t-k}^{z}) = \mathbb{V}(e_{i,t-k}^{z}) + \mathbb{CV}(\Delta^{k} z_{i,t} - \mathbb{E}[\Delta^{k} z_{i,t}|\Omega_t], e_{i,t-k}^{z}) - \mathbb{CV}(\mathbb{E}[z_{i,t-k}|\Omega_t] - \mathbb{E}[z_{i,t-k}|\Omega_{t-k}], e_{i,t-k}^{z}).$$

When k is small, the second and third terms are close to zero because there is little change in $z_{i,t}$ and the data releases between weeks $t - k$ and t only contain a small amount of incremental information concerning $z_{i,t-k}$. Consequently, the autocorrelations, $\rho_k \equiv \mathbb{CV}(e_{i,t}^{z}, e_{i,t-k}^{z})/\mathbb{V}(e_{i,t-k}^{z})$, will be close to one. As k rises, the sequence of data releases between weeks $t - k$ and t becomes more informative so that $\mathbb{E}[z_{i,t-k}|\Omega_t]$ approaches $z_{i,t-k}$. In fact when k reaches the reporting lag for variable i, $\mathbb{E}[z_{i,t-k}|\Omega_t] = z_{i,t-k}$, so ρ_k equals zero.[10] In sum, therefore, the autocorrelation patterns in Panel A of Table 2 reflect the relatively slow pace at which precise information concerning the underlying macro variable becomes available from the flow of data releases.

Panel B of Table 2 reports the cross-correlations between the real-time estimation errors for the six macro variables. As the table shows, the correlations are larger (in absolute value) between the US variables than between the German variables, but are all well below one. This feature of the data will allow me to examine whether order flows convey information on some macro variables rather than others.

One potential concern with the use of the real-time errors is that they utilize estimates of $\mathbb{E}[z_{i,\mathrm{M(m)}}|\Omega_{\mathfrak{d}}]$ and $\mathbb{E}[z_{i,\mathrm{Q}(i)}|\Omega_{\mathfrak{d}}]$ that are each constructed from a statistical model. If the models do a poor job of identifying these conditional expectations, the statistical power of the two-step procedure will be reduced. To alley this

---

[9] The real-time estimates for US variables use data releases on: quarterly GDP, Nonfarm Payroll, Employment, Retail Sales, Industrial Production, Capacity Utilization, Personal Income, Consumer Credit, Personal Consumption Expenditures, New Home Sales, Durable Goods Orders, Construction Spending, Factory Orders, Business Inventories, the Government Budget Deficit, the Trade Balance, NAPM index, Housing Starts, the Index of Leading Indicators, Consumer Prices and M1. The real-time estimates for German variables use data releases on GDP, Employment, Retail Sales, Industrial Production, Manufacturing Output, Manufacturing Orders, the Trade Balance, Consumer Prices and M1.

[10] Specifically, when $\mathbb{E}[z_{i,t-k}|\Omega_t] = z_{i,t-k}$,

$$\mathbb{CV}(e_{i,t}^{z}, e_{i,t-k}^{z}) = \mathbb{V}(e_{i,t-k}^{z}) + \mathbb{CV}(z_{i,t} - \mathbb{E}[z_{i,t}|\Omega_t], e_{i,t-k}^{z}) - \mathbb{CV}(z_{i,t-k} - \mathbb{E}[z_{i,t-k}|\Omega_{t-k}], e_{i,t-k}^{z}).$$

Clearly, the first and last term cancel, while the second equals zero because $e_{i,t-k}^{z} \in \Omega_t$.

	Mean	Max	Skewness	Autocorrelations			
	Std.	Min	Kurtosis	ρ_1	ρ_2	ρ_4	ρ_8
A:							
(i) US GDP	0.165	3.166	0.133	0.903	0.807	0.616	0.372
	1.341	-3.637	2.566				
(ii) US CPI	-0.064	0.379	0.265	0.749	0.528	0.528	0.520
	0.125	-0.369	3.196				
(iii) US M1	0.292	14.349	0.037	0.495	0.103	0.171	0.112
	3.921	-11.495	3.753				
(iv) German GDP	-1.255	8.406	0.001	0.922	0.843	0.701	0.387
	3.295	-11.742	4.412				
(v) German CPI	1.871	15.026	0.127	0.935	0.862	0.752	0.660
	5.536	-12.906	2.934				
(vi) German M1	-3.694	8.363	-1.288	0.795	0.585	0.393	0.237
	5.567	-29.020	7.284				
B: Cross-Correlations							
		(i)	(ii)	(iii)	(iv)	(v)	
(i) US GDP							
(ii) US CPI		-0.417*					
(iii) US M1		0.239*	-0.120*				
(iv) German GDP		0.100	-0.024	0.043			
(v) German CPI		-0.093	-0.109	0.043	0.105		
(vi) German M1		-0.098	-0.004	0.092	-0.049	-0.055	

C: Forecast Comparisons	M.M.S.		Real-Time	
	Mean	M.S.E	Mean	M.S.E
(i) US GDP	0.729	1.310	0.190	1.407
(ii) US CPI	-0.327	1.797	0.054	2.357
(iii) US M1	0.399	11.807	0.033	11.932
(iv) German GDP	0.132	6.981	-0.416	6.954
(v) German CPI	-0.136	1.687	-0.035	1.906
(vi) German M1	4.778	42.363	-0.159	20.561

Notes: The table reports summary statistics for the real-time estimation errors for US GDP, CPI and M1, and German GDP, CPI and M1 at the weekly frequency between January 1993 and June 1999. The real-time errors for the CPI and M1 are computed from (21) and for GDP from (22) using the data releases listed in footnote xx. The last four columns of Panel A report autocorrelations ρ_i at lag i. P-values for the null that $\rho_i = 0$ are all less than 0.01. Panel C compares the mean and Mean Squared Error (M.S.E.) of real-time estimates against the real-time errors computed from M.M.S. surveys or professional money managers. Statistical significance at the 5% level is denoted by *.

concern, Panel C of Table 2 compares the real-time estimates against the errors made by a set of professional money managers. On the Friday before each scheduled data release, M.M.S. surveys approximately forty money managers on their estimate for the upcoming release. Under the columns headed M.M.S., Panel C reports the mean and M.S.E. of the estimation error implied by the median response from the survey. The

mean and M.S.E of the real-time estimation errors using information, $\Omega_{\mathfrak{d}}$, available at the time of the survey are reported under the Real-Time columns. As the table shows, the real-time estimation errors are quite comparable to the M.M.S. survey errors. This finding provides assurance that the real-time estimation errors are not dominated by specification error.

2.3 Empirical Results

I begin by examining the relationship between excess returns and the six end-user order flows. This analysis serves three purposes. First, it is now well-established that order flows computed from interbank transactions account for a large percentage of the variation in spot rates over horizons ranging from hours to weeks. However, here I am concerned with order flow's role as a carrier of macro information over weeks and months. It is therefore important to quantify the relationship between order flows and excess returns over these longer horizons. Second, the order flow data I use comes from the transactions between end-users and one large bank rather than all the banks that act as market-makers in the spot USD/EUR market. Since data on market-wide end-user order flows is simply unavailable, we cannot directly examine how closely Citibank's flows proxy for market-wide order flow. We can, however, indirectly infer this from the strength of the relationship between excess returns and Citibank's flows. The strength of this relationship is also informative from a third perspective: It provides a benchmark for judging the degree to which spot rates respond to specific forms of macro information conveyed by order flow.

Table 3 documents the strength of the order flow/excess return relationship. Here I present the results of regressing excess returns between the start of weeks t and $t+\tau$, $er^{\tau}_{t+\tau} \equiv s_{t+\tau} - s_t + \hat{r}^{\tau}_t - r^{\tau}_t$ for $\tau = \{1, 4, 8\}$, on a constant, and the order flows from the six segments between the start of weeks $t-8$ and $t+\tau$.[11] The body of the table reports the OLS coefficients together with asymptotic standard errors (Newey and West 1987) corrected for heteroskedasticity and the presence of serial correlation in the form of an MA$(\tau - 1)$ process induced by the presence of overlapping forecasts.

The estimation results contain two noteworthy features. First, the coefficients on the individual order flow segments are quite different from each other. Some are positive, some are negative, some are highly statistically significant, others are not. In particular, the coefficients on US corporate and both US and non-US investor flows appear highly significant across specifications and horizons. By contrast, none of the coefficients on non-US corporate flows are significant, nor are the coefficients on the hedge fund flows when the corporate and investor flows are also included at the 4 and 8 week horizons.

To place these results in perspective, it is important to remember that: (i) the order flows are correlated across user types, and (ii) the estimated regression is not a structural equation for excess returns. The former observation means that no one regression coefficient summarizes the total impact of a single order flow on the foreign currency price dealers quote. For example, the negative coefficients on US investor flows do not signify that dealers viewed orders to purchase euros by this group as a signal that the euro was

---

[11] One week euro-currency rates were unavailable over the entire sample period so the results in Panel A use a proxy for excess returns using four week rates instead. Re-estimating the regressions in the Panel A with one week rates over the sub-sample where they are available gives very similar results. The results are also essentially unchanged when I use $s_{t+\tau} - s_t$ as the regressand and add the interest differential, $r^{\tau}_t - \hat{r}^{\tau}_t$, as a regressor. In this case, the estimated coefficients on $r^{\tau}_t - \hat{r}^{\tau}_t$ have a negative sign but are not statistically significant, while those on the order flows are very similar to the values reported in Table 3. The results are also very similar if the order flows are cumulated starting in weeks $t-12$ or $t-4$.

Horizon	Corporate		Hedge		Investor		R^2	χ^2
	US	Non-US	US	Non-US	US	Non-US		(p-value)
A: 1 Week								
	0.617**	-0.069					0.030	10.052
	(0.211)	(0.087)						(0.007)
			0.183**	-0.019			0.027	8.560
			(0.063)	(0.131)				(0.014)
					-0.410**	0.254**	0.052	16.621
					(0.165)	(0.082)		(<0.001)
	0.409*	-0.020	0.126*	-0.117	-0.372**	0.211**	0.081	24.708
	(0.221)	(0.092)	(0.067)	(0.133)	(0.168)	(0.085)		(<0.001)
B: 4 Weeks								
	0.581**	-0.038					0.124	21.625
	(0.134)	(0.053)						(<0.001)
			0.111**	0.034**			0.054	7.776
			(0.042)	(0.090)				(0.020)
					-0.376**	0.234**	0.234	37.088
					(0.100)	(0.046)		(<0.001)
	0.384**	-0.021	0.051	-0.066	-0.346**	0.203**	0.308	53.832
	(0.114)	(0.048)	(0.037)	(0.081)	(0.104)	(0.045)		(<0.001)
C: 8 Weeks								
	0.485**	-0.009					0.195	30.497
	(0.093)	(0.040)						(<0.001)
			0.072**	0.014			0.047	4.304
			(0.035)	(0.074)				(0.116)
					-0.266**	0.175**	0.310	46.766
					(0.083)	(0.030)		(<0.001)
	0.356**	-0.007	0.011	-0.044	-0.227**	0.157**	0.419	83.702
	(0.088)	(0.034)	(0.027)	(0.058)	(0.079)	(0.028)		(<0.001)

Notes: The table reports coefficients and standard errors from regressions of excess returns between the start of weeks t and $t + \tau$, $er^{\tau}_{t+\tau}$ for $\tau = \{1, 4, 8\}$ on a constant (estimates not reported), and order flows cumulated from the start of weeks $t - 8$ and $t + \tau$. Estimated coefficients on the order flows are multiplied by 1000. The right hand column reports χ^2 statistics for the null that all the coefficients on order flows are zero. Estimates are calculated at the weekly frequency. The standard errors correct for heteroskedasticity and the MA$(\tau - 1)$ error process. Statistical significance at the 5% and 1% level is denoted by * and **.

overvalued. The second observation follows from the fact that we are looking at the end-user flows received by a single bank. Currency trading models such as Lyons (1997) and Evans and Lyons (1999) predict that dealers revise their price quotes in response to aggregate order flow between dealers, which in turn reflects the trading strategies of individual dealers that are driven by the end-user flows they receive at their respective banks. The micro-based model presented in Section 2 abstracted from this mapping between individual end-user flows and aggregate interdealer flows, but we should not overlook this simplification when interpreting the results in Table 3. The coefficients on the individual end-user flows received by Citibank should be

statistically significant insofar as they proxy for the unobserved market-wide order flow that is driving quote revision, but the individual parameters do not have any simple structural interpretation.

The second noteworthy feature of the results in Table 3 concerns the strength of the relation between excess returns and order flow. The right hand columns of the table show that the order flows are jointly significant at conventional levels in all but one case. Furthermore, the R^2 statistics indicate that the proportion of the variation in excess returns accounted for by all six flows rises from approximately 8 to 41 percent as the horizon increases from one to eight weeks. It is instructive to compare this pattern of explanatory power against other results in the literature. For example, Evans and Lyons (2002a & 2002b) found that aggregate interdealer order flows accounted for between 45 and 78 percent of the variation in spot rate returns at the daily frequency, figures far higher than we see in Table 3. Similarly, in a very large data sample, Berger et al. (2005) found that the explanatory power of interdealer order flows is between 40 and 50 percent at frequencies ranging from a few minutes to a week. When judged against these findings, the results in Table 3 do not suggest that Citibank's six flows provide a particularly accurate proxy for aggregate interdealer order flow at the weekly horizon. By contrast, the explanatory power of the Citibank's flows at the four and eight week horizons is somewhat higher than results reported elsewhere using aggregate interdealer order flow. In particular, Berger et al. (2005) find that aggregate order flow accounts for approximately 20 percent of the variation in spot rate returns at the one and two-month horizons, whereas Table 3 shows that Citibank's flows account for between 30 and 40 percent. The point here is not to suggest that Citibank's order flows are a perfect proxy for market-wide order flow, they are surely not. However, the strength of the relationship between excess returns and the available flow data is sufficiently strong at the four and eight week horizons to provide useful information about the role order flow plays in conveying information to the market as a whole, rather than just to Citibank.

I now turn to the central question: Does order flow convey new price-relevant information to dealers concerning the current state of the macroeconomy? To address this question, I first examine whether the six end-user flows convey new macro information. Table 4 reports the results of regressing the real-time estimation errors for macro variable $z_{i,t}$ at the start of week t, $e_{i,t}^z = z_{i,t} - \mathbb{E}[z_{i,t}|\Omega_t]$, on the six end-user flows, $x_{j,t}$:

$$e_{i,t}^z = \sum_{j=1}^{6} b_j x_{j,t} + v_t \tag{23}$$

In these regressions, the end-user flows are cumulated between the start of weeks $t-8$ and either $t+4$ or the week before data on $z_{i,t}$ is released, whichever period is shorter.[12]

If the Citibank's end-user flows convey no information about $z_{i,t}$ beyond that contained in current and past data releases (i.e. the elements of Ω_t), the b_j coefficients in regression (23) should be close to zero and insignificant. The body of Table 4 reports OLS estimates of b_j together with asymptotic standard errors corrected for heteroskedasticity. Two features of the results stand out. First, the estimated b_j coefficients are jointly significant at the one percent level for every macro variable considered. Furthermore, the coefficient on each end-user flow is significant at the one percent level for at least one of the macro variables. By this metric,

---

[12] The flows are cumulated between start of weeks $t-8$ and $t+4$ in the case of the real-time GDP estimation errors because the reporting lags for GDP are always longer than four weeks. For prices and money the reporting lags are typically two weeks, so the flows are cumulated between start of weeks $t-8$ and $t+2$ at the end of each month. There is no quantitative difference in the estimation results if all the flows are cumulated between the start of weeks $t-8$ and $t+2$.

Table 4: Real-Time Estimation Errors and Order Flows

| Real-Time Error | Corporate | | Hedge | | Investor | | R^2 | χ^2 |
	US	Non-US	US	Non-US	US	Non-US		(p-value)
A: US								
GDP	-0.530**	0.010	0.133**	0.109	0.428**	-0.256**	0.197	89.430
	(0.137)	(0.059)	(0.049)	(0.098)	(0.100)	(0.043)		(<0.001)
CPI	0.296	0.252**	-0.112**	-0.153	-0.572**	0.255**	0.157	197.056
	(0.181)	(0.054)	(0.048)	(0.098)	(0.107)	(0.046)		(<0.001)
M1	-0.243	-0.090	0.052	0.178*	0.255**	-0.242**	0.128	54.024
	(0.133)	(0.061)	(0.042)	(0.089)	(0.118)	(0.051)		(<0.001)
B: Germany								
GDP	0.106	0.100	0.120**	-0.147	-0.092	-0.065	0.029	19.873
	(0.175)	(0.064)	(0.058)	(0.093)	(0.143)	(0.052)		(0.003)
CPI	-0.380**	-0.188**	0.048	0.045	-0.131	-0.068	0.018	33.917
	(0.144)	(0.049)	(0.047)	(0.109)	(0.106)	(0.048)		(<0.001)
M1	1.081**	0.146**	-0.122**	-0.043	0.101	0.182**	0.145	96.927
	(0.242)	(0.057)	(0.055)	(0.132)	(0.125)	(0.048)		(<0.001)
C: Difference								
GDP	0.636**	0.090	-0.013	-0.256**	-0.520**	0.191**	0.068	59.258
	(0.213)	(0.092)	(0.071)	(0.126)	(0 .159)	(0.060)		(<0.001)
CPI	-0.676**	-0.440**	0.160**	0.198	0.441**	-0.324**	0.082	131.419
	(0.247)	(0.067)	(0.069)	(0.163)	(0.162)	(0.078)		(<0.001)
M1	1.324**	0.237**	-0.174**	-0.221	-0.154	0.424**	0.163	149.297
	(0.256)	(0.077)	(0.065)	(0.150)	(0.158)	(0.073)		(<0.001)

Notes: The table reports coefficients and standard errors from regression (23). The estimated coefficients on the order flows are multiplied by 1000. The right hand column reports χ^2 statistics for the null that all the coefficients on order flows are zero. Estimates are calculated at the weekly frequency. The standard errors correct for heteroskedasticity. Statistical significance at the 5% and 1% level is denoted by * and **.

every end-user flow conveys some incremental macro information. The second noteworthy feature concerns the level of explanatory power. In Panel A, the R^2 statistics range from 13 to 20 percent. These results indicated that the flows jointly contain an economically significant amount of incremental information about the three US macro variables. In the case of the German variables, the picture is more mixed. Panel B shows that the flows only account for between two and three percent of the variations in the real-time errors for GDP and the CPI, but over 14 percent in the case of M1. Insofar as spot rates reflect the difference between US and EU monetary policy, order flows may carry more information about the difference in macroeconomic conditions between countries. To investigate this possibility, Panel C reports results for cases where the difference between the German and US real-time errors is the regressand in (23). The estimation results lie somewhere between those for the US and German variables for GDP and the CPI, but in case of M1, the

R^2 statistic is somewhat higher. In sum, it appears that the flows carry significant incremental information on both US GDP and CPI and the difference between the German and US M1.

The results in Table 4 provide direct empirical evidence in support of the idea that order flows convey more timely information about the state of the macroeconomy to dealers than is available from the flow of information contained in macro data releases. By construction, the real-time estimation errors are a function of $z_{i,t}$ and the data releases in Ω_t; they are not derived from any financial data on prices, interest rates or order flows. Thus, the results in Table 4 cannot be attributable to a spurious correlation between some form of measurement error in the real-time estimate and Citibank's order flows. Further, recall that the real-time errors are comparable to the errors implied by the M.M.S. survey of professional money managers, so it is hard to argue that they embody very inefficient estimates, $\mathbb{E}[z_{i,t}|\Omega_t]$. Instead, the rather striking implication of the results in Table 4 is that Citibank's end-user order flows carry timely information about current macroeconomic conditions.[13]

Is the macro information carried by Citibank's end-user flows price-relevant? Recall from the model that dealers revise their spot rate quotes in response to new information on current macroeconomic conditions insofar as it changes their view about the future course of short-term interest rates and the aggregate hedging demand for foreign currency. Clearly, information on GDP, the CPI and M1 could qualify on both counts. What is less clear is whether the macro information available to Citibank is also transmitted to other banks (either by their own end-user flows or by Citibank's interdealer trading) in a sufficiently transparent form that it is used to revise the spot rate quotes of dealers across the market.

To address this issue, I estimate regressions of the form

$$er_{t+\tau}^{\tau} = \lambda_0 + \sum_{i=1}^{k} \lambda_i \hat{E}[e_{i,t}^z|x_t] + v_{t+\tau}, \tag{24}$$

where $\hat{E}[e_{i,t}^z|x_t]$ is the fitted value from regression (23) for a set of k macro variables. If there is sufficient transparency in the spot market for the macro information found in Citibank's flows to be used by dealers when revising their spot price quotes, the estimates of λ_i should be significant. Moreover, if the estimated projections $\hat{E}[e_{i,t}^z|x_t]$ for the k macro variables well approximate the complete flow of macro information dealers use to revise their quotes, movements in $\hat{E}[e_{i,t}^z|x_t]$ should account for the contribution of order flows to the variation in excess returns.

Estimation of (24) is complicated by the fact that the regressors, $\hat{E}[e_{i,t}^z|x_t]$, are themselves estimates, and so contain sampling variation. To account for this, I replace $\hat{E}[e_{i,t}^z|x_t]$ with the real-time errors, $e_{i,t}^z$, and then estimate the λ_i parameters using $\hat{E}[e_{i,t}^z|x_t]$ as instruments for each $e_{i,t}^z$. This instrumental variable procedure is akin to 2SLS and provides identical λ_i estimates as OLS. I compute standard errors for the λ_i parameters from the instrumental variable procedure with the Newey and West (1987) covariance estimator that allows for the presence of heteroskedasticity and a $MA(\tau-1)$ error process.

---

[13] The results in Table 4 are most closely related to the findings of Rime, Sarno, and Sojli (2007). These authors consider regressions of $e_{i,t}^z$ computed from survey estimates for individual data releases, z_i, on aggregate interdealer order flow in the USD/EUR, USD/GBP and USD/JPY markets over a twelve month sample starting in February 2004. They find that the coefficients on at least one of the order flows are significant at the 10 percent level for every US, EU and UK data release they study. These findings are consistent with the results in Table 4, but they have less statistical precision because they are based on a much shorted data sample.

Table 5: Excess Returns and Real-Time Estimation Errors

Horizon	GDP German	GDP US	CPI German	CPI US	M1 German	M1 US	R^2
A: 4 Weeks							
	0.142 (0.109)	-0.681** (0.103)					0.098
			0.500** (0.202)	0.514** (0.112)			0.112
					0.117 (0.069)	-0.704** (0.137)	0.185
	0.137 (1.904)	-0.110 (0.801)	-1.140 (1.452)	0.480 (0.838)	1.647 (2.284)	0.690 (1.151)	0.287
B: 4 Weeks Differences							
	0.427** (0.077)						0.147
			-0.173** (0.038)				0.044
					0.264** (0.042)		0.113
	0.895** (0.429)		0.517** (0.246)		0.191 (0.304)		0.232
C: 8 Weeks							
	0.177** (0.082)	-0.581** (0.076)					0.316
Differences	0.394** (0.057)						0.242

Notes: The table reports coefficients and standard errors from regression (24). The estimated coefficients on the order flows are multiplied by 1000 and are estimated using the projections $\mathbb{E}[e_{i,t}^z|\Omega_t]$ as instruments for the real-time errors $e_{i,t}^z$. Estimates are calculated at the weekly frequency. The standard errors correct for heteroskedasticity and an MA($\tau - 1$) error process. Statistical significance at the 5% and 1% level is denoted by * and **.

Panel A of Table 5 shows the results from estimating (24) with excess returns computed over a four week horizon. The estimates in the first three rows indicate the order flows do indeed convey price-relevant information about GDP, the CPI and M1. The λ_i coefficients for the German CPI, US GDP, CPI and M1 are all significant at the one percent level. However, as the forth row of the table shows, the λ_i coefficients are no longer individually significant when all six projections are included. Panel B reports estimates of (24) where the estimated projections of the difference between the German and US real-time errors replace $\hat{E}[e_{i,t}^z|x_t]$. Recall from Table 4 that the order flows have significant explanatory power for each of these differences, and so appear to carry timely information about the cross-country differences in each macro variable. In Panel B of Table 5 we see strong evidence that this information is price-relevant. The λ_i coefficients on the GPD and CPI projections are highly significant. Moreover, the estimated projections jointly account for approximately 75 percent of order flows explanatory power of excess returns: Table 3 showed that the R^2

statistic for the regression of excess returns on the six end-user flows is 31 percent at the four week horizon, while the comparable statistic in Panel B of Table 5 is 23 percent.

The results in Panel C are equally striking. Here I report the estimates of the λ_i coefficients for cases where excess returns are computed over an eight week horizon and projections of the real-time GDP errors are used as regressors. These projections are computed in the same manner as those shown in Table 4 except the order flows are cumulated from week $t-8$ to either $t+8$ or the week before the GDP data release. As the table shows, the estimated λ_i coefficients on the projections for the German and US real-time errors are highly significant with opposite signs. Furthermore, the R^2 statistic of 32 percent implies that the two projections account for approximately 75 percent of order flows explanatory power of excess returns at the eight week horizon.

The results in Table 5 also point to an interesting pattern between the flows of macro information on GDP and the revision of dealers' spot rate quotes. Throughout the table, the λ_i estimates show that dealers revise their quotes upwards when the end-user flows imply that their prior estimates of current German GDP are too low, and their prior estimates of current US GDP are too high. These findings map closely into the structure of the model.[14] Recall that dealers expect the future interest differential between euro and dollar rates to be higher when their estimates of the difference between the EU and US output gap widens. So, if dealers receive new information via order flow that current US GDP is higher than previously estimated, they will lower their spot rate quotes in anticipation that the FED will follow a tighter monetary policy that will lower the future path of the interest differential between euro and dollar rates. Similarly, if dealers learn from order flow that their prior estimates of current German GDP are too low, they will revise their spot rate quotes upwards in anticipation of tighter monetary policy by the ECB. It is also interesting to note that the estimated coefficients on the US GDP error projections are three to four times the absolute size of their counterparts for German GDP. It appears that dealers react more strongly to new information concerning US GDP than German GDP. This difference may reflect the fact that the end-user flows carry more incremental information concerning US GDP than German GDP, as shown in Table 4. It could also arise because dealers anticipate that the FED and ECB will react differently to future macroeconomic conditions.

3   Conclusion

In this paper I have sought to reconnect the high frequency dynamics of spot exchange rates with developments in the macroeconomy. To do so, I first presented a new type of exchange-rate model that bridges the gap between existing microstructure and macro models. It combines the dealer-centered view of spot rate determination from currency trading models with forecasts for interest rates consistent with their role as the policy instruments of central banks – a feature emphasized in recent macro models. In so doing, I provide a richer picture of the possible links between spot rates, order flow and macroeconomic developments than is found in existing exchange-rate models. In particular, the model shows that order flow can connect

---

[14]Unfortunately, the pattern of estimated coefficients on the projections of the real-time CPI errors do not have a similar model-based interpretation. The estimates change sign and significance depending on the specification. These results may reflect the fact that information concerning current prices has ambiguous implications for dealers' forecasts of future monetary policy. The information could lead dealers to revise their forecasts for future inflation and the real exchange rate – factors that have offsetting implications for future monetary policy.

excess returns with macroeconomic developments if it conveys price-relevant macro information to dealers who have limited information about the current state of the economy. I then examine this implication empirically using order flows from six groups of end-users, and the real-time estimation errors for six macro variables. My results show that the end-user flows carry a significant amount of timely information about current macroeconomic conditions. Furthermore, this information appears price-relevant in the sense that it is reflected in how dealers revise their spot rate quotes across the market.

Finally, let me offer a wider perspective on my results. While the explanatory power of the projections for excess returns in Table 5 are very high relative to other models, they account for less than half the variance. The model therefore falls well short of identifying all the information dealers are using to revise their spot rate quotes. What could be the nature of this missing information? One possibility is that the microeconomic environments driving the current demand for foreign currency are correlated with future macroeconomic conditions. Under these circumstances, order flow conveys information to dealers about the future course of interest rates as well as current macro conditions, so the estimated flows of macro information examined here understate the true information flow dealers are using. Incorporating these effects into an estimable structural model for the high frequency dynamics for spot rates using order flows and macro variables should provide further progress in resolving the exchange-rate disconnect puzzle.

References

Bacchetta, P. and E. van Wincoop (2006, June). Can information heterogeneity explain the exchange rate determination puzzle? *American Economic Review 96*(3), 552–576.

Bacchetta, P. and E. van Wincoop (2008, June). Higher order expectations in asset pricing. *Journal of Money, Credit and Banking* (3).

Berger, D. W., A. P. Chaboud, S. V. Chernenko, E. Howorka, R. S. K. Iyer, D. Liu, and J. H. Wright (2005). Order flow and exchange rate dynamics in electronic brokerage system data. International Finance Discussion Papers 830, Board of Governors of the Federal Reserve System.

Clarida, R., J. Gal, and M. Gertler (1998). Monetary policy rules in practice-New Research in Macroeconomics, Chapter 2. *European Economic Review 42*(6), 1033–1067.

Dominguez, K. M. and F. Panthaki (2006, February). What defines news in foreign exchange markets? *Journal of International Money and Finance 25*(1), 168–198.

Engel, C., N. Mark, and K. West (2007). Exchange Rate Models are Not as Bad as You Think. *NBER Working Paper*.

Engel, C. and K. West (2006). Taylor Rules and the Deutschmark–Dollar Real Exchange Rate. *Journal of Money, Credit, and Banking 38*(5).

Engel, C. and K. D. West (2005, June). Exchange rates and fundamentals. *Journal of Political Economy 113*(3), 485–517. available at http://ideas.repec.org/a/ucp/jpolec/v113y2005i3p485-517.html.

Evans, M. D. D. (2005, December). Where are we know? real-time estimates of the macroeconomy. *International Journal of Central Banking 1*(6), 127–175.

Evans, M. D. D. and V. Hnatkovska (2007). International financial integration and the real economy. *IMF Staff Papers 54*, 220–269.

Evans, M. D. D. and R. K. Lyons (1999). Order flow and exchange rate dynamics. *NBER Working Paper*.

Evans, M. D. D. and R. K. Lyons (2002a). Informational integration and fx trading. *Journal of International Money and Finance 21*(6), 807–831.

Evans, M. D. D. and R. K. Lyons (2002b). Order flow and exchange rate dynamics. *Journal of Political Economy 110*(1), 170–180.

Evans, M. D. D. and R. K. Lyons (2004, March). A new micro model of exchange rate dynamics. Working Paper 10379, National Bureau of Economic Research.

Evans, M. D. D. and R. K. Lyons (2005). Do currency markets absorb news quickly? *Journal of International Money and Finance 24*(6), 197–217.

Evans, M. D. D. and R. K. Lyons (2007, June). Exchange rate fundamentals and order flow. Working Paper 13151, National Bureau of Economic Research.

Evans, M. D. D. and R. K. Lyons (2008a, April). How is macro news transmitted to exchange rates? *Journal of Financial Economics 88*(1), 26–50.

Evans, M. D. D. and R. K. Lyons (2008b, June). Macroeconomics and exchange rates dynamics redux. Working paper, Department of Economics, Georgetown Unversity.

Froot, K. A. and T. Ramadorai (2005). Currency returns, intrinsic value, and institutional-investor flows. *Journal of Finance 60*(3), 1535–1566.

Love, R. and R. Payne (2007). Macroeconomic news, order flows and exchange rates. *Journal of Financial and Quantitative Analysis*. Forthcoming.

Lyons, R. K. (1995, October). Tests of microstructural hypothesis in the foreign exchange market. *Journal of Financial Economics 39*, 321–351.

Lyons, R. K. (1997, May). A simultaneous trade model of the foreign exchange hot potato. *Journal of International Economics 42*, 275–298.

Mark, N. (2005). Changing Monetary Policy Rules, Learning, and Real Exchange Rate Dynamics. *NBER Working Paper*.

Meese, R. and K. Rogoff (1983). Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample? *Journal of International Economics 14*(1-2), 3–24.

Newey, W. and K. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Obstfeld, M. and K. Rogoff (2001). The Six Major Puzzles in International Macroeconomics: Is There a Common Cause? *Nber Macroeconomics Annual 2000*.

Osler, C. L. (2008). *Springer Encyclopedia of Complexity and System Science.*, Chapter Foreign Exchange Microstructure: A Survey.

Rime, D., L. Sarno, and E. Sojli (2007, April). Exchange rate forecasting, order flow and macroeconomic information. Working Paper 2, Norges Bank.

Sager, M. J. and M. P. Taylor (2007). Commercially available order flow data and exchange rate movements: *Caveat Emptor. Journal of Money, Credit and Banking.* Forthcoming.

A   Appendix (Not for Publication)

I provide a brief description of how I computed the real-time estimates of a monthly log series $\varkappa$. Computing the real-time estimates for a quarterly series like GDP follows analogously and is described in detail by Evans (2005). Let $\Delta\varkappa_d$ denote the daily increment to the monthly value for $\varkappa_{\mathrm{M}(\tau)}$, where $\mathrm{M}(\tau)$ is the last day of month τ. Next, define the partial sum

$$\widetilde{\Delta\varkappa_d} \equiv \sum_{i=\mathrm{M}(\tau-1)+1}^{\min\{\mathrm{M}(\tau),d\}} \Delta\varkappa_i,$$

as the cumulative daily contribution to $\varkappa_{\mathrm{M}(\tau)}$ in month τ. Notice that when $d = \mathrm{M}(\tau)$, the monthly change in $\varkappa_{\mathrm{M}(\tau)}$, $\Delta^{\mathrm{M}}\varkappa_{\mathrm{M}(\tau)} = \widetilde{\Delta\varkappa_{\mathrm{M}(\tau)}}$. The daily dynamics of $\widetilde{\Delta\varkappa_d}$ are described by

$$\widetilde{\Delta\varkappa_d} = (1 - \mathfrak{D}_d)\,\widetilde{\Delta\varkappa_{d-1}} + \Delta\varkappa_d, \tag{A1}$$

where $\mathfrak{D}_d$ is a dummy variable equal to one on the first day of each month, and zero otherwise. To accommodates the presence of variable reporting lags, let $\Delta^{\mathrm{M}(j)}\varkappa_d$ denote the monthly growth in $\varkappa$ ending on day $\mathrm{M}(\tau - j)$ where $\mathrm{M}(\tau)$ denotes the last day of the most recently completed month and $d \geq \mathrm{M}(\tau)$. Monthly growth in the last (completed) month is given by

$$\Delta^{\mathrm{M}(1)}\varkappa_d = (1 - \mathfrak{D}_d)\,\Delta^{\mathrm{M}(1)}\varkappa_{d-1} + \mathfrak{D}_d\widetilde{\Delta\varkappa_{d-1}}. \tag{A2}$$

When d is the first day of a new month, $\mathfrak{D}_d = 1$, so $\Delta^{\mathrm{M}(1)}\varkappa_{\mathrm{M}(\tau)+1} = \widetilde{\Delta\varkappa_{\mathrm{M}(\tau)}} = \Delta^{\mathrm{M}}\varkappa_{\mathrm{M}(\tau)}$. On all other days, $\Delta^{\mathrm{M}(1)}\varkappa_d = \Delta^{\mathrm{M}(1)}\varkappa_{d-1}$. To accommodate occasions where the reporting lag is more than a month, I track monthly growth two months back via the recursion:

$$\Delta^{\mathrm{M}(2)}\varkappa_d = (1 - \mathfrak{D}_d)\,\Delta^{\mathrm{M}(2)}\varkappa_{d-1} + \mathfrak{D}_d\Delta^{\mathrm{M}(1)}x_{d-1}. \tag{A3}$$

Equations (A1), (A2) and (A3) enable me to define the link between the daily contributions, $\Delta\varkappa_d$, and data releases. Suppose the reporting lag for the release on day d is less than one month. Then if $\widehat{\Delta\varkappa_d}$ is the released value for the growth in $\varkappa$ during the last month on day t,

$$\widehat{\Delta\varkappa_d} = \Delta^{\mathrm{M}(1)}\varkappa_d. \tag{A4}$$

If the reporting lag is longer than a month (but less than two),

$$\widehat{\Delta\varkappa_d} = \Delta^{\mathrm{M}(2)}\varkappa_d. \tag{A5}$$

I incorporate the information contained in the monthly data releases on other variables in a similar manner. (Incorporating information from quarterly data releases is more complex, see Evans 2005 for details.) Specifically, let z_d^i denote the value of another series, released on day d, that relates to activity in the last completed month. I assume that

$$z_d^i = \beta_i\Delta^{\mathrm{M}(1)}\varkappa_d + u_d^i. \tag{A6}$$

where u_d^i is an i.i.d. $N(0, \sigma_i^2)$ shock. In cases where the reporting lag is two months,

$$z_d^i = \beta_i \Delta^{\mathrm{M}(2)} \varkappa_d + u_d^i. \tag{A7}$$

It is important to recognize that (A4) - (A7) allow for variations in the reporting lag from data release to data release.

To complete the model, I specify the dynamics for the daily increments, $\Delta \varkappa_d$. I assume that

$$\Delta \varkappa_d = \sum_{i=1}^{\varrho} \phi_i \Delta^{\mathrm{M}(i)} \varkappa_d + \zeta_d^{\varkappa}, \tag{A8}$$

where $\zeta_d^{\varkappa}$ is an i.i.d. $N(0, \sigma_{\zeta}^2)$ shock.

Finding the real-time estimates of $\varkappa$ requires a solution to two related problems. First, there is a pure inference problem of how to compute $\mathbb{E}[\varkappa_{\mathrm{M}(\tau)} | \Omega_d]$ using the signalling equations (A4) - (A7), and the $\Delta \varkappa_d$ process in (A8), given values for all the parameters in these equations. Second, I need to estimate these parameters. The Kalman Filtering algorithm provides a solution to both problems. In particular, given a set of parameter values, the algorithm provides the means to compute the real-time estimates $\mathbb{E}[\varkappa_{\mathrm{M}(\tau)} | \Omega_d]$. The algorithm also allows me to construct a sample likelihood function from the data series, so that the model's parameters can be computed by maximum likelihood.

To use the algorithm, I write the model in state space form. For the case where $\varrho = 1$, the dynamics described by equations (A1) - (A3) and (A8) can be represented by the matrix equation:

$$
\begin{bmatrix}
\widetilde{\Delta \varkappa_d} \\
\Delta^{\mathrm{M}(1)} \varkappa_d \\
\Delta^{\mathrm{M}(2)} \varkappa_d \\
\Delta \varkappa_d
\end{bmatrix}
=
\begin{bmatrix}
1 - \mathfrak{D}_d & 0 & 0 & 1 \\
\mathfrak{D}_d & 1 - \mathfrak{D}_d & 0 & 0 \\
0 & \mathfrak{D}_d & 1 - \mathfrak{D}_d & 0 \\
0 & \phi_1 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
\widetilde{\Delta \varkappa_{d-1}} \\
\Delta^{\mathrm{M}(1)} \varkappa_{d-1} \\
\Delta^{\mathrm{M}(2)} \varkappa_{d-1} \\
\Delta \varkappa_{d-1}
\end{bmatrix}
+
\begin{bmatrix}
0 \\
0 \\
0 \\
\zeta_d^{\varkappa}
\end{bmatrix},
$$

or, more compactly

$$\mathbb{Z}_d = \mathbb{A}_d \mathbb{Z}_{d-1} + \mathbb{V}_d. \tag{A9}$$

This is the state equation of the state space form.

The link between the data releases on $\varkappa$ and the elements of $\mathbb{Z}_t$ are described by (A4) and (A5):

$$\widehat{\Delta \varkappa_d} = \begin{bmatrix} 0 & \mathrm{ML}_d^1(\varkappa) & \mathrm{ML}_d^2(\varkappa) & 0 \end{bmatrix} \mathbb{Z}_d, \tag{A10}$$

where $\mathrm{ML}_d^i(\varkappa)$ denotes a dummy variable that takes the value of one when the reporting lag for series $\varkappa$ lies between $i-1$ and i months, and zero otherwise. The link between the releases for the other series and elements of $\mathbb{Z}_d$ are described by (A6) and (A7):

$$z_d^i = \begin{bmatrix} 0 & \beta_i \mathrm{ML}_d^1(z^i) & \beta_i \mathrm{ML}_d^2(z^i) & 0 \end{bmatrix} \mathbb{Z}_d + u_d^i. \tag{A11}$$

A2

Stacking (A10) and (A11) for series $i = 1, 2..., g$ gives

$$
\begin{bmatrix} \widehat{\Delta \varkappa_d} \\ z_d^1 \\ \vdots \\ z_d^g \end{bmatrix} = \begin{bmatrix} 0 & \mathrm{ML}_d^1(\hat{\varkappa}) & \mathrm{ML}_d^2(\hat{\varkappa}) & 0 \\ 0 & \beta_1 \mathrm{ML}_d^1(z^1) & \beta_1 \mathrm{ML}_d^2(z^1) & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \beta_g \mathrm{ML}_d^1(z^g) & \beta_g \mathrm{ML}_d^2(z^g) & 0 \end{bmatrix} \mathbb{Z}_d + \begin{bmatrix} 0 \\ u_d^1 \\ \vdots \\ u_d^g \end{bmatrix},
$$

or

$$
\mathbb{X}_d = \mathbb{C}_d \mathbb{Z}_d + \mathbb{U}_d. \tag{A12}
$$

This equation links the vector of *potential* data releases for day d, $\mathbb{X}_d$, to elements of $\mathbb{Z}_d$. The vector of actual data releases for day d, $\mathbb{Y}_d$, is related to the vector of *potential* releases by

$$
\mathbb{Y}_d = \mathbb{B}_d \mathbb{X}_d,
$$

where $\mathbb{B}_d$ is a $n \times (g + 1)$ selection matrix that "picks out" the $n \geq 1$ data releases for day d. Combining this expression with (A12) gives me the observation equation:

$$
\mathbb{Y}_d = \mathbb{B}_d \mathbb{C}_d \mathbb{Z}_d + \mathbb{B}_d \mathbb{U}_d. \tag{A13}
$$

Equations (A9) and (A13) describe a state space form which can be used to find real-time estimates of variable $\varkappa$ in two steps. In the first, I obtain the maximum likelihood estimates of the model's parameters. For this purpose the sample likelihood function is built up recursively by applying the Kalman Filter to (A9) and (A13). The second step applies the Kalman Filter to (A9) and (A13) to calculate the real-time estimates of $\varkappa$ using the maximum likelihood parameter estimates.

The real-time estimates for US variables use data releases on quarterly GDP and 18 monthly releases: Nonfarm Payroll, Employment, Retail Sales, Industrial Production, Capacity Utilization, Personal Income, Consumer Credit, Personal Consumption Expenditures, New Home Sales, Durable Goods Orders, Construction Spending, Factory Orders, Business Inventories, the Government Budget Deficit, the Trade Balance, NAPM index, Housing Starts, the Index of Leading Indicators, Consumer Prices and M1. The real-time estimates for German variables use data releases on quarterly GDP and 8 monthly releases: Employment, Retail Sales, Industrial Production, Manufacturing Output, Manufacturing Orders, the Trade Balance, Consumer Prices and M1. I allow for 10 lags in the daily increment process when estimating real-time GDP, and 7 lags for the other variables. These specifications appear to capture all the time-series variation in the data. In particular, I am unable to reject the null hypothesis of no serial correlation in the Kalman Filter innovations evaluated at the maximum likelihood estimates for any of our models.