A ROBUST NON-LINEAR MULTIVARIATE KALMAN FILTER FOR ARBITRAGE IDENTIFICATION IN HIGH FREQUENCY DATA

P. J. BOLLAND AND J. T. CONNOR
London Business School
Department of Decision Science
Sussex Place, Regents Park
London NW1 4SA
Phone -(+44) 71-262-5050
FAX - (+44) 71-724-7875
E-Mail - pbolland@lbs.lon.ac.uk
E-Mail - jconnor@lbs.lon.ac.uk

ABSTRACT

We present a methodology for modelling real world high frequency financial data. The methodology copes with the erratic arrival of data and is robust to additive outliers in the data set. Arbitrage pricing relationships are formulated into a linear state space representation. Arbitrage opportunities violate these pricing relationships and are analogous to multivariate additive outliers. Robust identification/filtering of arbitrage opportunities in the data is accomplished by Kalman filtering. The state space model used to describe the pricing relationships is general enough to handle both linear and non-linear models. The recursive Kalman equations are adapted to filter tick data, cope with the erratic arrival of observations and produce estimates of all the arbitrage prices on every time step. We demonstrate the methodology with a robust neural network filter applied to foreign exchange triangular arbitrage. Tick data from three markets is used: $/DM, £/$, £/DM 1993-1995. The filter produces estimates of the arbitrage price for all exchange rates on every second, increasing both the speed and efficiency of arbitrage identification.

KEYWORDS: Arbitrage, Foreign Exchange, Multivariate Kalman Filter, Neural Network, Outliers, Robust, Tick Data.

1. Introduction

Arbitrage is a fundamental mechanism for achieving efficiency in the financial markets (Ross 1976). An arbitrage opportunity occurs when a price discrepancy exists between two or more highly related assets. The opportunity can be exploited by buying the under priced asset and selling the over priced asset, producing a profit without incurring any risk. Mispricing is rapidly corrected in highly competitive markets (Frenkel and Levich 1975,1977), therefore arbitrage traders need rapid identification, fast transactions and low transaction costs. Many arbitrage relationships have been identified in the financial markets. Our methodology can be applied to any system of linear arbitrage pricing relationships. Section 1.1 describes the triangular foreign exchange arbitrage we use to demonstrate the methodology. Previous studies of arbitrage identification have mainly been limited to examining daily data and so might miss many of available intraday opportunities. Studies that have examined intraday data (Rhee and Chang 1992) have been limited to examining only a minute fraction of the data because of the need for simultaneous observations. The methodology we present allows arbitrage opportunities to be identified with irregular (non-simultaneous) observations.

Irregular times series presents a serious problem to conventional modelling methodologies. Several methodologies for dealing with erratic data have been suggested in the literature. Muller et al 1990, suggest methods of linear interpolation between erratic observations to obtain a regular homogenous times series. Other authors (Ghysels and Jasiak 1995) have favoured non-linear time deformation ("business-time" or "tick-time"), however this methodology has no simple equivalent for multivariate series. The methodology we present describes the dynamics of fundamental underlying arbitrage states which are observed as erratic noisy exchange rates. We treat the erratic arrival as a missing data problem. The Kalman filter described is discrete, as the data is only provide in quantised time steps (i.e. seconds), however the methodology could be extended to continuous time problems with the Kalman-Bucy filter (Meditch 1969). The state space representation described in section 2 allows us model the system at the maximum resolution of the available data (Reuters data quoted by the second) .

Conventional modelling methodologies may also be inappropriate for modelling tick data as the distribution is often heavy tailed (Dacorogna 1995). Financial data, especially quotations, are prone to data corruption and outliers. Chung 1991, discovered 0.25% of the MMI futures quotes were outside of the daily high and low and are therefore serious data corruption's. Section 3 details our robust methodology which is similar to that described by Masreliez and Martin 1975 and Martin and Vandaele 1983.

The state space representation is capable of incorporating both linear and non-linear models. The estimation of the models is performed using an E.M. algorithm described in section 4, which was introduced by Dempster, Laird and Rubin 1977 to estimate parameters of models when some of the data is missing.

The methodology we present is suitable for real world financial data and increases both the speed and efficiency of arbitrage identification. We demonstrate the filter on $/£, $/DM, £/DM data from 1993-1995, the results of the estimation and filtering are shown in section 5.

1.1 Foreign Exchange Arbitrage

We examine foreign exchange triangulation's for arbitrage opportunities (although the same methodology can be applied to many varieties of arbitrage). In the absence of transactions costs and bid-ask spread the following equilibrium relationships must hold for currency rates,

$$\begin{aligned}
&\mathrm{EX}(0,1)\ \mathrm{EX}(1,2)\ \mathrm{EX}(2,0) = 1 \\
&\mathrm{EX}(0,1)\ \mathrm{EX}(1,2)\ \mathrm{EX}(2,3)\ \mathrm{EX}(3,0) = 1 \\
&\vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\
&\mathrm{EX}(0,1)\ \mathrm{EX}(1,2)\ \mathrm{EX}(2,3)\ldots \quad \ldots \mathrm{EX}(m,0) = 1,
\end{aligned} \tag{1}$$

where EX(i,j) represents the spot rate for currency j when expressed in units of currency i. If the equilibrium relationships in Eq.(1) hold, then a single countries *m* exchange rates can be used to produce estimates of all the cross rates, $\mathrm{EX}(i, j) = \mathrm{EX}(0, j) / \mathrm{EX}(0, i)$, in this paper the US Dollar is used as the base currency. Taking logarithms of the triangular relationships, allows the cross rates to be expressed as

$$\begin{aligned}
&\log(\mathrm{EX}(1,2)) = \log(\mathrm{EX}(0,2)) \ - \ \log(\mathrm{EX}(0,1)) \\
&\quad \vdots \qquad\qquad\qquad \vdots \qquad\qquad\qquad \vdots \\
&\log(\mathrm{EX}(i, j)) = \log(\mathrm{EX}(0, j)) - \log(\mathrm{EX}(0, i)) \\
&\quad \vdots \qquad\qquad\qquad \vdots \qquad\qquad\qquad \vdots \\
&\log(\mathrm{EX}(m-1, m)) = \log(\mathrm{EX}(0, m)) \ - \ \log(\mathrm{EX}(0, m-1)).
\end{aligned} \tag{2}$$

If the additive relationships of Eq.(2) are violated, an arbitrage opportunity exists where riskless, profitable transactions can occur. Violations of the triangular relationships are analogous to an outlier in the data set, the larger the mispricing the larger the outlier. When market friction' s are included (transaction costs and bid-ask spread) slight mispricing can occur within small bands around the arbitrage price. In the following section the triangular currency relationships are encoded within a state space form and a multivariate Kalman filter is used to identify any significant violations of Eq.(2).

2. Space Representation of FX Relationships

The methodology we present below describes how outliers can be robustly identified/filtered in multivariate non-linear data. In this application the outliers that the Kalman filter identifies are situations in which an arbitrage opportunities exist. The Kalman filter has been adapted to filter tick data and to update the estimates of the exchange rates every time step. The Kalman filter used is general enough to handle both linear and non-linear models. For non-linear models a point-wise linearization is performed to predict the Kalman filter's state changes, and to update the recursive estimates of the error prediction covariance (Connor, Martin, Atlas 1994). The parameters of the models used in the Kalman filter are robustly estimated from cleaned data, described in section 4.

The observation vector $\mathbf{z}_t$ in the state space model represents the logarithm of each exchange rate observed. If all possible exchange rates (ticks) are observed in a given second then, $\mathbf{z}_t$ is given by

$$\mathbf{z}_t = (z_t^{(0,1)}, z_t^{(0,2)}, \ldots, z_t^{(0,m)}, z_t^{(1,2)}, \ldots, z_t^{(1,m)}, \ldots, z_t^{(m-1,m)})^T, \qquad (3)$$

where $z_t^{(i,j)} = \log(\mathrm{EX(i, j)})$. Usually only a subset of Eq.(3) are observed. The elements of $\mathbf{z}_t$ come in two principle groups :

- The log of the *m* exchange rates for the base currency (0,j),

$$z_t^{(0,1)}, z_t^{(0,2)}, \ldots, z_t^{(0,m)} \qquad = \mathrm{Log}\,(\textit{Base Rates})\,. \qquad (4)$$

- The log of the corresponding cross rates (i,j),

$$z_t^{(1,2)}, \ldots, z_t^{(1,m)}, z_t^{(2,3)}, \ldots, z_t^{(m-1,m)} = \mathrm{Log}\,(\textit{Cross Rates})\,. \qquad (5)$$

The exchange rate mispricing problem can be formulated into a familiar state space model,

$$\mathbf{x}_t = \mathbf{f}(\mathbf{x}_{t-1}) + \mathbf{e}_t \;, \qquad (6)$$

$$\mathbf{z}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{v}_t \;. \qquad (7)$$

The state vector $\mathbf{x}_t$ in Eq.(6) represents the log of the arbitrage value of the *m* exchange rates for the base currency ($) as well as the auto-regressive structure of the system. The state transition vector $\mathbf{f}(\mathbf{x}_{t-1})$ in Eq.(6) represents the system dynamics that may be linear or non-linear (in the case of a linear system $\mathbf{f}(\mathbf{x}_{t-1})$ is simply the state transition matrix Φ_t). The observation matrix, $\mathbf{H}_t$ in Eq.(7), extracts the base rates and uses the logarithmic arbitrage equations to estimates the cross rates The system described in Eq.(6) and Eq.(7) have two types of driving noise, $\mathbf{e}_t$ the state noise and $\mathbf{v}_t$ the observation noise. The state noise $\mathbf{e}_t$ represents the variation due to the exchange rates underlying arbitrage dynamics. The observation noise $\mathbf{v}_t$ has two components $\mathbf{v}_t = \mathbf{u}_t + \mathbf{w}_t$, the first component $\mathbf{u}_t$, represents the variation caused by the transaction costs and bid ask spread, which allow the price to move freely within unprofitable bounds. The second component $\mathbf{w}_t$, represents the additive outliers within the data (whether they are data corruption's or market misprice anomalies). The state transition vector $\mathbf{f}(\mathbf{x}_{t-1})$ in Eq.(6), can be described as a non-linear multivariate auto-regressive (NMAR) process for each of the base currency's exchange rates. The multivariate auto-regressive process, NMAR($p^{(1)}, p^{(2)}, \ldots, p^{(m)}$) is defined by,

$$
\begin{aligned}
x_t^{(1)} &= f^{(1)}(x_{t-1}^{(1)},\ldots,x_{t-p^{(1)}}^{(1)},x_{t-1}^{(2)},\ldots,x_{t-p^{(2)}}^{(2)},\ldots,x_{t-1}^{(m)},\ldots,x_{t-p^{(m)}}^{(m)}) + \varepsilon_t^{(1)} \\
x_t^{(2)} &= f^{(2)}(x_{t-1}^{(1)},\ldots,x_{t-p^{(1)}}^{(1)},x_{t-1}^{(2)},\ldots,x_{t-p^{(2)}}^{(2)},\ldots,x_{t-1}^{(m)},\ldots,x_{t-p^{(m)}}^{(m)}) + \varepsilon_t^{(2)} \\
\vdots \quad & \qquad \vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\quad \vdots \\
x_t^{(m)} &= f^{(m)}(x_{t-1}^{(1)},\ldots,x_{t-p^{(1)}}^{(1)},x_{t-1}^{(2)},\ldots,x_{t-p^{(2)}}^{(2)},\ldots,x_{t-1}^{(m)},\ldots,x_{t-p^{(m)}}^{(m)}) + \varepsilon_t^{(m)},
\end{aligned}
\tag{8}
$$

where $x_t^{(1)}, x_t^{(2)}, \ldots, x_t^{(m)}$ are the log exchange rates for the base currency, and $p^{(i)}$ is the number of autoregressive terms for the i th exchange rate. In Eq.(8) $f^{(i)}$'s denote non-linear functions governing each individual exchange rate. The state vector $\mathbf{x}_t$, the state transition vector $\mathbf{f}(\mathbf{x}_{t-1})$ and the disturbance vector $\mathbf{e}_t$ in Eq.(6) are defined as follows:

$$
\mathbf{x}_t = (x_t^{(1)},\ldots,x_{t-p^{(1)}+1}^{(1)},x_t^{(2)},\ldots,x_{t-p^{(2)}+1}^{(2)},\ldots,x_t^{(m)},\ldots,x_{t-p^{(m)}+1}^{(m)})^T, \tag{9}
$$

$$
\begin{aligned}
\mathbf{f}(\mathbf{x}_{t-1}) = (& f^{(1)}(\mathbf{x}_{t-1}), x_{t-1}^{(1)},\ldots,x_{t-p^{(1)}+1}^{(1)}, \\
& f^{(2)}(\mathbf{x}_{t-1}), x_{t-1}^{(2)},\ldots,x_{t-p^{(2)}+1}^{(2)}, \quad \ldots, \\
& \ldots, \quad f^{(m)}(\mathbf{x}_{t-1}), x_{t-1}^{(m)},\ldots,x_{t-p^{(m)}+1}^{(m)})^T,
\end{aligned}
\tag{10}
$$

$$\mathbf{e}_t = (\varepsilon_t^{(1)}, 0, \ldots, \varepsilon_t^{(2)}, 0, \ldots, \varepsilon_t^{(m)}, 0, \ldots)^T , \tag{11}$$

where $\varepsilon_t^{(j)}$, the random state noise associated with the j th exchange rate, appears in the $1 + \sum_{i=1}^{j} p^{(i-1)}$ position in the disturbance vector $\mathbf{e}_t$ (where $p^{(0)} = 0$).

The observation matrix $\mathbf{H}_t$ in Eq.(7) extracts the base rates and the cross rates from the state vector $\mathbf{x}_t$. Each of the rows of $\mathbf{H}_t$ relate to a specific exchange rate, the rows are defined as follows,

$$\mathbf{H}_t = [\mathbf{h}_t^{(0,1)\mathrm{T}}, \quad \cdots, \quad \mathbf{h}_t^{(0,m)\mathrm{T}}, \quad \mathbf{h}_t^{(1,2)\mathrm{T}}, \quad \cdots, \quad \mathbf{h}_t^{(1,m)\mathrm{T}}, \quad \cdots, \quad \mathbf{h}_t^{(m-1,m)\mathrm{T}}]. \tag{12}$$

For *base currency exchange rate* (0,j) :

$$\mathbf{h}_t^{(0,j)} = [0 \quad \mathbf{0}_{1x\,p^{(1)}} \quad 0 \quad \ldots \quad \mathbf{0}_{1x\,p^{(j-1)}} \quad 1 \quad \mathbf{0}_{1x\,p^{(j)}} \quad 0 \quad \ldots \quad 0 \quad \mathbf{0}_{1x\,p^{(m)}}], \tag{13}$$

so each $\mathbf{h}_t^{(0,j)}$ extracts the base currency rate x_t^j from $\mathbf{x}_t$.

For *cross currency exchange rates* (i,j) :

$$\mathbf{h}_t^{(i,j)} = [0 \quad \mathbf{0}_{1x\,p^{(1)}} \quad 0 \quad \ldots \quad \mathbf{0}_{1x\,p^{(i)}} \quad -1 \quad \mathbf{0}_{1x\,p^{(j)}} \quad 1 \quad \ldots \quad 0 \quad \mathbf{0}_{1x\,p^{(m)}}], \tag{14}$$

so each $\mathbf{h}_t^{(i,j)}$ estimates the cross rate (i,j) using the additive log relationships (i.e. $\mathbf{h}_t^{(i,j)} \cdot \mathbf{x}_t = x_t^j - x_t^i$).

For regular (evenly spaced) time series all the rates would be observed on every time step. Tick data, however, requires a methodology capable of modelling irregular time series. On any given second only currencies for which a trade (tick) has occurred enter the observation vector and only the rows of the observation matrix which correspond to an actual observation are used to update the filtering equations. The observation noise vector $\mathbf{v}_t$, in Eq.(7) also reconfigures it's dimensionality to correspond to the actual trades that occur. This gives rise to several possible situations :

- No observations on any market,

$$\mathbf{z}_t = [NULL], \qquad \mathbf{H}_t = [NULL], \qquad \mathbf{v}_t = [NULL].$$

The observation vector $\mathbf{z}_t$, the observation matrix $\mathbf{H}_t$ and the observation noise $\mathbf{v}_t$ are set to null.

- One or more markets produce observations,
  e.g. base rate $z_t^{(0,i)}$ and cross rate $z_t^{(j,k)}$ are traded.

$$\mathbf{z}_t = \begin{bmatrix} z_t^{(0,i)} \\ z_t^{(j,k)} \end{bmatrix}, \qquad \mathbf{H}_t = \begin{bmatrix} \mathbf{h}_t^{(0,i)} \\ \mathbf{h}_t^{(j,k)} \end{bmatrix}, \qquad \mathbf{v}_t = \begin{bmatrix} v_t^{(0,i)} \\ v_t^{(j,k)} \end{bmatrix}.$$

- All the markets produce observations in one second,

$$\mathbf{z}_t = \begin{bmatrix} z_t^{(0,1)} \\ \vdots \\ z_t^{(0,m)} \\ z_t^{(1,2)} \\ \vdots \\ z_t^{(1,m)} \\ \vdots \\ z_t^{(m-1,m)} \end{bmatrix}, \qquad \mathbf{H}_t = \begin{bmatrix} \mathbf{h}_t^{(0,1)} \\ \vdots \\ \mathbf{h}_t^{(0,m)} \\ \mathbf{h}_t^{(1,2)} \\ \vdots \\ \mathbf{h}_t^{(1,m)} \\ \vdots \\ \mathbf{h}_t^{(m-1,m)} \end{bmatrix}, \qquad \mathbf{v}_t = \begin{bmatrix} v_t^{(0,1)} \\ \vdots \\ v_t^{(0,m)} \\ v_t^{(1,2)} \\ \vdots \\ v_t^{(1,m)} \\ \vdots \\ v_t^{(m-1,m)} \end{bmatrix}.$$

Expanding and contracting the observation equation in this way, allows the state space model to cope with the erratic arrival of tick data and immediately incorporate all new information to update the state estimates for all exchange rates. The methodology produces an estimate of the states (exchange rates), an estimate of the associated prediction error covariance, as well as the predictions of future states at every second, regardless of any tick being observed.

3. Robust Kalman Filter

The underlying states $\mathbf{x}_t$ in Eq.(6), are unknown, they are estimated by a robust Kalman filter (Kalman 1961). Using robust methodologies protects the modelling procedure from serious performance degradation caused by ill conditioned data. The recursive non-Gaussian Kalman filter equations as described by Masreliez 1975 and Martin and Vandaele 1983 are detailed below. The robust one step ahead predicted state vector $\hat{\mathbf{x}}_t$ and the predicted observation vector $\hat{\mathbf{z}}_t$ are given by,

$$\hat{\mathbf{x}}_t = \mathbf{f}(\tilde{\mathbf{x}}_{t-1}), \tag{15}$$
$$\hat{\mathbf{z}}_t = \mathbf{H}_t \hat{\mathbf{x}}_t. \tag{16}$$

where $\tilde{\mathbf{x}}_{t-1}$ is the filtered state vector. The distribution terms $\mathbf{e}_t$ and $\mathbf{v}_t$ in Eq.(6) and Eq.(7) are assumed to be zero mean, serially independent and mutually independent, however no assumptions about their distributions are made. The covariance matrices of $\mathbf{e}_t$ and $\mathbf{v}_t$ are denoted by $\mathbf{Q}_t = E(\mathbf{e}_t \cdot \mathbf{e}_t')$ and $\mathbf{R}_t = E(\mathbf{v}_t \cdot \mathbf{v}_t')$ respectively. The modelling methodology we employ assumes that the noise covariance matrices remain constant over time. For financial data this assumption may be invalid (Ruiz 1994). The methodology we present can be extend to incorporate stochastic volatility, see Harvey Ruiz and Shephard 1992. In an effort to limit the impact of stochastic volatility the estimation of $\mathbf{Q}_t$ and $\mathbf{R}_t$ was made using a rolling window, see section 4. The robustly filtered estimate of the state vector $\tilde{\mathbf{x}}_{t-1}$ and the prediction error covariance matrix $\mathbf{M}_t$ are defined by the following recursive relationships,

$$\tilde{\mathbf{x}}_t = \hat{\mathbf{x}}_t + \mathbf{M}_t \mathbf{H}_t \mathbf{g}_t(\bar{\mathbf{z}}_t)\,, \tag{17}$$

$$\mathbf{M}_{t+1} = \Phi_t \mathbf{P}_t \Phi_t' + \mathbf{Q}_t\,, \tag{18}$$

$$\mathbf{P}_t = \mathbf{M}_t - \mathbf{M}_t \mathbf{H}_t' \mathbf{G}_t(\bar{\mathbf{z}}_t) \mathbf{H}_t \mathbf{M}_t\,, \tag{19}$$

where $\bar{\mathbf{z}}_t = \mathbf{z}_t - \hat{\mathbf{z}}_t$ is the innovations vector (the observed residual), $\mathbf{g}_t(\bar{\mathbf{z}}_t)$ is the score function of the innovations with components,

$$\{\mathbf{g}_t(\bar{\mathbf{z}}_t)\}_i = -\left[\frac{\partial\, p\{\bar{\mathbf{z}}_t | Z_{t-1}\}}{\partial(\bar{\mathbf{z}}_t)_i}\right] \cdot [p\{\bar{\mathbf{z}}_t | Z_{t-1}\}]^{-1}\,, \tag{20}$$

and $\mathbf{G}_t(\bar{\mathbf{z}}_t)$ is defined as the differential of the score function, with elements,

$$\{G_t(\bar{\mathbf{z}}_t)\}_{ij} = \frac{\partial\{\mathbf{g}_t(\bar{\mathbf{z}}_t)\}_i}{\partial(\bar{\mathbf{z}}_t)_j}\,. \tag{21}$$

For a non-linear state space model the state transition matrix Φ_t in Eq.(18) is estimated by a point-wise linearization of the non-linear model. The elements of Φ_t are the partial derivatives of $\mathbf{f}$ evaluated about the robustly filtered estimates of the state vector $\tilde{\mathbf{x}}_t$,

$$\Phi_{i,j} = \left.\frac{\partial f(\mathbf{x}_t)_i}{\partial x_j}\right|_{\mathbf{x}_t = \tilde{\mathbf{x}}_t}\,. \tag{22}$$

In the standard Kalman filter the density function for the innovations is assumed to be Gaussian. In order to obtain robustness we assume $\bar{\mathbf{z}}_t$ has a symmetric heavy-tailed density function. The score function $\mathbf{g}_t(\bar{\mathbf{z}}_t)$ for a Gaussian innovation process is linear. For a heavy tailed density $\mathbf{g}_t(\bar{\mathbf{z}}_t)$ is given by a non-linear gain function that limits the influence of large innovations. There is some latitude given in choice of $\mathbf{g}_t(\bar{\mathbf{z}}_t)$ in the above equations. The score functions for the Gaussian distribution, Huber's least informative distribution (Huber 1981) and the Hampel re-descending function are shown in figure 1 (for the one-dimensional case).

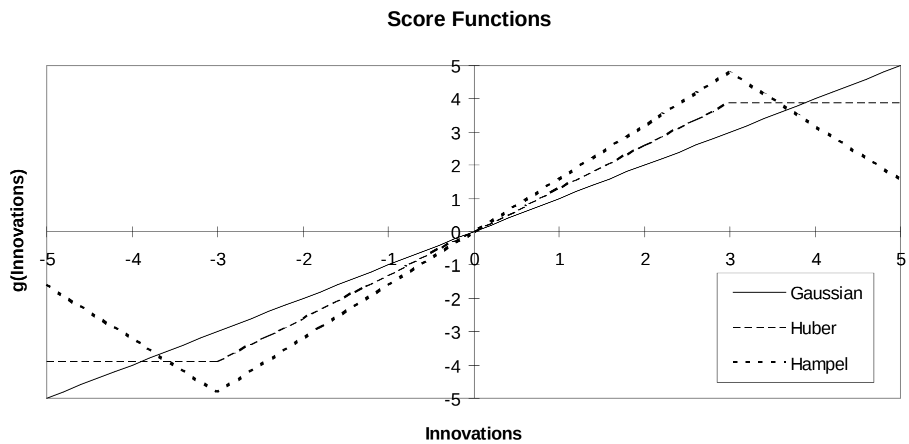

Figure 1: Score Functions: Gaussian, Huber, Hampel.

For the case of Gaussian innovations the score function $\mathbf{g}_t(\bar{\mathbf{z}}_t)$ is given by,

$$\mathbf{g}_t(\bar{\mathbf{z}}_t) = (\mathbf{H}_t\mathbf{M}_t\mathbf{H}_t' + \mathbf{R}_t)^{-1} \cdot \bar{\mathbf{z}}_t \,. \tag{23}$$

The derivations of $\mathbf{g}_t(\bar{\mathbf{z}}_t)$ for the n-dimensional Huber and Hampel densities is given in Bolland and Connor 1995.

The size of the innovations is the critical value which determines whether the observation is an outlier (arbitrage opportunity). The magnitude of the outlier is defined by $r_t^2 = (\mathbf{z}_t - \hat{\mathbf{z}}_t)' \Sigma^{-1}(\mathbf{z}_t - \hat{\mathbf{z}}_t)$ where Σ is the covariance matrix of the innovations. The measurement r_t allows us to set a definition for an outlier, so that only mispricing of sufficient magnitude to allow for profitable trades are identified.

4. Model Specification and Estimation

In order to produce the robust Kalman filter, estimates of $\mathbf{f}(\mathbf{x}_{t-1})$, $\mathbf{Q}_t$ and $\mathbf{R}_t$ need to be obtained. The state transition function $\mathbf{f}(\mathbf{x}_{t-1})$ can be approximated with many different modelling methodologies. There is a large body of empirical evidence to suggest that the dominant structure in $\mathbf{f}(\mathbf{x}_{t-1})$ will be a mean reverting process (Fama 1965). The mean reversion could be an artificial artefact of the data. Roll 1984, showed that bid-ask bounce induced strong negative auto-correlation in financial data. Time series of market prices contain both bid and ask prices so if no new information arrives the true value remains constant, any observed variation is caused by the difference in bid and ask price. Bouncing between bid and ask prices gives rise to a strong negative auto-correlation shown in figure 2.

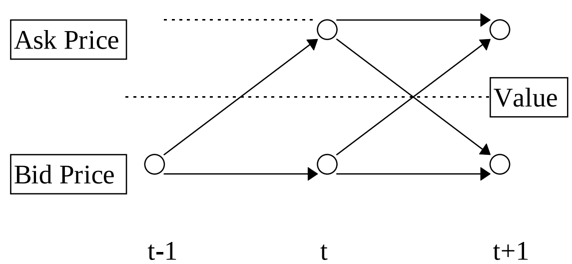

Figure 2: Bid-Ask Bounce.

To limit the impact of the bid-ask bounce, mid-prices were modelled. The mean reversion could also be an artefact of the price quantisation (prices quoted in discrete units).

The states of the system described are the arbitrage values of the exchange rates. The first differences of the state were taken to produce a stationary series. Predicting state changes rather than there levels requires a slight re-formulation of the $f_t^i(\mathbf{x}_{t-1})$'s in Eq.(10). The state transitions are formed by two components, a random walk component, $x_{t-1}^{(i)}$ (the previous state), plus the state changes $d^{(i)}(\mathbf{x}_{t-1} - \mathbf{x}_{t-2})$, so Eq.(10) becomes,

$$
\begin{aligned}
\mathbf{f}(\mathbf{x}_{t-1}) = (x_{t-1}^{(1)} + d^{(1)}(\mathbf{x}_{t-1} - \mathbf{x}_{t-2}), x_{t-1}^{(1)}, \dots, x_{t-p^{(1)}+1}^{(1)},& \\
x_{t-1}^{(2)} + d^{(2)}(\mathbf{x}_{t-1} - \mathbf{x}_{t-2}), x_{t-1}^{(2)}, \dots, x_{t-p^{(2)}+1}^{(2)}, \quad \dots,& \qquad (24) \\
\dots, \quad x_{t-1}^{(m)} + d^{(m)}(\mathbf{x}_{t-1} - \mathbf{x}_{t-2}), x_{t-1}^{(m)}, \dots, x_{t-p^{(m)}+1}^{(m)})^T,&
\end{aligned}
$$

where $d^{(i)}(\mathbf{x}_{t-1} - \mathbf{x}_{t-2})$ represents the NMAR structure of the state changes.

Tick data for three currencies \$/DM, £/\$, £/DM (1993-1995) was used to demonstrate the methodology. The fundamental states are therefore the arbitrage values of \$/DM, £/\$, so $\mathbf{x}_t = (\log(\$ / \mathrm{DM}^*), \log(\$ / £^*))^t$. The states were estimated by several different models:

- Naive random walk model (no filter), $\tilde{\mathbf{x}}_t = \tilde{\mathbf{x}}_{t-1}$
- Random walk Kalman filter, $\tilde{\mathbf{x}}_t = \tilde{\mathbf{x}}_{t-1} + \mathbf{M}_t \mathbf{H}_t \mathbf{g}_t(\bar{\mathbf{z}}_t)$
- Linear MAR(1,1), $\tilde{\mathbf{x}}_t = f(\tilde{\mathbf{x}}_{t-1}) + \mathbf{M}_t \mathbf{H}_t \mathbf{g}_t(\bar{\mathbf{z}}_t)$
- Neural Network NMAR(n,n), $\tilde{\mathbf{x}}_t = f(\tilde{\mathbf{x}}_{t-1}, \ldots, \tilde{\mathbf{x}}_{t-n}) + \mathbf{M}_t \mathbf{H}_t \mathbf{g}_t(\bar{\mathbf{z}}_t)$

To produce an estimation set for an MAR(n,n) we require data examples with n consecutive state changes across all states. Due to the erratic nature of the tick data constructing an estimation data set without missing observations for with more than one lag proved difficult. An initial linear MAR(1,1) model was estimated from the data where ticks occurred on consecutive seconds for both the £/\$ and £/DM exchange rates. Using the MAR(1,1) as an estimate of $\mathbf{f}(\mathbf{x}_{t-1})$, the robust Kalman filter was applied to the whole data set produced a filtered data set. This data set contained the filtered values of the actual observed state changes and the estimated state changes produced by the simple linear model. An MAR(n,n) estimation set could now be constructed from the filtered data, with the estimated state changes filling the problematic missing observations. To avoid the problems of non-stationarity in the dynamics over the whole data set (two years), the parameters were estimated on a rolling window of one trading week, and re-estimated daily.

4.1 Estimation of Model Parameters

An estimation maximisation (EM) algorithm is employed to estimate the neural network parameters, denoted λ, of Eq.(10) or equivalently Eq.(24). The EM algorithm, see Dempster, Laird, and Rubin (1977), is the standard approach when estimating model parameters with missing data. The EM algorithm has been used in the neural network community before, see for example Jordan and Jacobs, and Connor, Martin, and Atlas (1994).

The E-Step, given in section 4.1, estimates the missing data. With the estimated missing data assumed to be true, the parameters are then chosen by way of maximising the likelihood. This procedure is iterative with new parameter estimates giving rise to new estimates of missing data which in turn give rise newer parameter estimates.

4.11 E-Step

During the estimation step, the missing data, namely the $\mathbf{x}_t$, $\mathbf{v}_t$, and $\mathbf{e}_t$, of Eq.(6) and Eq.(7) must be estimated. This is accomplished using the robust Kalman filter of section 3. The estimated missing data is denoted $\tilde{\mathbf{x}}_t$, $\tilde{\mathbf{v}}_t$, and $\tilde{\mathbf{e}}_t$.

4.12 M-Step

The robust likelihood for the system defined by Eq.(6) and Eq.(7) is given by

$$l(\lambda) = \prod_{t=1}^{N} p\{\bar{\mathbf{z}}_t | \mathbf{Z}_{t-1}\} \qquad (25),$$

where $p\{\bar{\mathbf{z}}_t | \mathbf{Z}_{t-1}\}$ is a function of r_t^2, the magnitude of the innovations $r_t^2 = -(\tilde{\mathbf{z}}_j - \hat{\mathbf{z}}_j(\mathbf{Z}_{t-1},\lambda))' \Sigma_j^{-1} (\tilde{\mathbf{z}}_j - \hat{\mathbf{z}}_j(\mathbf{Z}_{t-1},\lambda))$ and $\Sigma_t = \mathbf{H}_t{}^T \mathbf{M}_t \mathbf{H}_t + \mathbf{R}_t$ is the covariance matrix. This has been derived by De Jong 1988 for the case where initial state estimates and noise variances are considered. The log likelihood is defined by

$$L(\lambda) = \sum_{j=1}^{N} \log(p\{\bar{\mathbf{z}}_t | \mathbf{Z}_{t-1}\}) \qquad (26)$$

The parameters to maximise the likelihood,

$$\hat{\lambda}^j = \underset{\lambda}{\arg\max} \sum_{t=1}^{N} \log(p\{\bar{\mathbf{z}}_t | \mathbf{Z}_{t-1}\}). \qquad (27)$$

Simple gradient descent is used to maximise (27), this is done by ignoring the dependence of Σ_t^{-1} on λ

$$\frac{\partial L_t(\lambda)}{\partial \lambda_i} = \sum_{t=1}^{N} \frac{\partial \log(p\{\bar{\mathbf{z}}_t | \mathbf{Z}_{t-1}\})}{\partial\, \mathbf{z}_t} \mathbf{H}_t \frac{\partial\, \hat{\mathbf{x}}_t(\tilde{\mathbf{x}}_{t-1},\cdots,\tilde{\mathbf{x}}_{t-p},\lambda)}{\partial \lambda_i} \qquad (28)$$

Noting that $\{\mathbf{g}(\mathbf{z}_t)\}_i = \dfrac{\partial \log(p\{\bar{\mathbf{z}}_t | \mathbf{Z}_{t-1}\})}{\partial\, (\mathbf{z}_t)_i}$ from Eq.(20) and using Eq.(17),

$$\mathbf{g}(\mathbf{z}_t)' \mathbf{H}_t \mathbf{M}_t = \left(\tilde{\mathbf{x}}_t - \hat{\mathbf{x}}_t(\tilde{\mathbf{x}}_{t-1},\cdots,\tilde{\mathbf{x}}_{t-p},\lambda)\right)' \qquad (29)$$

which can be used with Eq.(28) to get the gradient in terms of clean data

$$\frac{\partial L_t(\lambda)}{\partial \lambda_i} = \sum_{t=1}^{N} \left(\tilde{\mathbf{x}}_t - \hat{\mathbf{x}}_t(\tilde{\mathbf{x}}_{t-1},\cdots,\tilde{\mathbf{x}}_{t-p},\lambda)\right)' \mathbf{M}_t^{-1} \frac{\partial \hat{\mathbf{x}}_t(\tilde{\mathbf{x}}_{t-1},\cdots,\tilde{\mathbf{x}}_{t-p},\lambda)}{\partial \lambda_i} \qquad (30)$$

which is equivalent to doing back-propagation on clean data.

Outlier Observations

When an outlier, $\mathbf{z_t}$, is observed r_t^2 will be very large causing the term $g.\left(r_t{}^2\right)$ to limit the contribution of $\mathbf{z}_t$ to the likelihood in (26). As mentioned in section 3, the prediction variance $\sum_t$ reflecting greater uncertainty in the fundamental rates. Future contributions to the likelihood will be effected by this increased uncertainty.

Univariate Case

If only a single quote is observed at a given time t, the associated gradient is given by

$$\frac{\partial\ L_t(\lambda)}{\partial\lambda_i} = -2(z_t - \hat{z}_t(\lambda))' s_{zz\,t}{}^{-1} \frac{\partial\ \hat{z}_t(\lambda)}{\partial\lambda} \qquad (31))$$

which if z_t corresponds to a fundamental dollar denominated rate is the same as found in the univariate case first explored in Connor, Martin and Atlas (1994).

Cross Currency

If the quote corresponds to a cross currency, related to the dollar denominated rates by $z_t^{(i,j)} = x_t^{(i)} - x_t^{(j)}$, the gradient will be composed of contributions from all the predictors related through $\mathbf{H}_t\mathbf{M}_t^{-1}$

$$\frac{\partial\ L_t(\lambda)}{\partial\lambda_i} = -2(z_t^{(i,j)} - \hat{z}_t^{(i,j)}(\lambda))' \sum_k (M_{k,i,t}^{-1} - M_{k,j,t}^{-1}) \frac{\partial\ \hat{x}_t^{(k)}(\lambda)}{\partial\lambda}. \qquad (32)$$

Several Quotes of the Same Currency

If several quotes are observed for one of the underlying dollar denominated rates, the learning algorithm simplifies greatly. The prediction covariance matrix and corresponding inverse are given by:

$$\sum_t = \begin{bmatrix} a & b & \cdots & b & b \\ b & a & \cdots & b & b \\ \vdots & \vdots & & \vdots & \vdots \\ b & b & \cdots & a & b \\ b & b & \cdots & b & a \end{bmatrix} \quad \sum_t{}^{-1} = \begin{bmatrix} c & d & \cdots & d & d \\ d & c & \cdots & d & d \\ \vdots & \vdots & & \vdots & \vdots \\ d & d & \cdots & c & d \\ d & d & \cdots & d & c \end{bmatrix}, \qquad (33)$$

where *c* and *d* are derived from $ac + (N_t - 1)bd = 1$ and $ad + bc + (N_t - 2)bd = 0$ which allows the reduction of Eq.(33) to the simpler form

$$\frac{\partial\ L_t(\lambda)}{\partial\lambda_i} = -2(\left(\frac{1}{N_t}\sum_{i=1}^{N_t} x^{(i)}{}_t\right) - \hat{x}_t(\lambda))' k_{N_t} \frac{\partial\ \hat{x}_t(\lambda)}{\partial\lambda} \qquad (34)$$

where $k_{N_t} = c + (N_t - 1)d$. Since several quotes are available, the additive noise is smoothed out and one has more confidence in the average of the quotes, $\frac{1}{N_t}\sum_{i=1}^{N_t} x^{(i)}{}_t$, than any of the quotes would be given alone. This added confidence is expressed in terms of the stronger gradient in (i) where $k_{N_t} > k_1$ for $N_t > 1$.

No Observations
Due to the erratic nature of tick data, often there will be no observations during a given period. For this missing data, there will be no contribution to the likelihood given in (26). But as in the case with extreme outliers, the uncertainty in future predictions, Σ_t will grow and effect the likelihood of future observations.

4.2 Estimation of Noise Variances

An iterative procedure was applied to produce estimates of $\mathbf{Q}_t$ and $\mathbf{R}_t$. The observation noise covariance $\mathbf{R}_t$, has two components, $\mathbf{u}_t$ the market friction's (bid-ask spread, transaction costs etc.) and $\mathbf{w}_t$ the additive outliers (pricing anomalies). As filtering procedure interpolates through additive outliers, the estimate of $\mathbf{R}_t$ is only dependent on the first component $\mathbf{u}_t$, $\mathbf{R}_t^* = E(\mathbf{u}_t \cdot \mathbf{u}_t')$. Initial estimate of observation noise covariance $\mathbf{R}_t^*$, and the state noise covariance $\mathbf{Q}_t$, were produced by maximising the likelihood of the MAR(1,1) state space model. These estimates of $\mathbf{R}_t^*$ and $\mathbf{Q}_t$ were refined during the neural network estimation process by repeated application of maximum likelihood.

5. Results

The tick data was obtained from Reuters, April 1993-April 1995, the mid-price was modelled. The initial linear MAR(1,1) model produced had a state change transitions matrix given by,

$$\mathbf{f}(\mathbf{x}_{t-1}) = \begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \end{bmatrix} + \begin{bmatrix} -0.5065 & 0.0358 \\ -0.0564 & -0.4361 \end{bmatrix} \cdot \begin{bmatrix} x_{t-1}^{(1)} - x_{t-2}^{(1)} \\ x_{t-1}^{(2)} - x_{t-2}^{(2)} \end{bmatrix}. \qquad (35)$$

The initial data set from which the MAR(1,1) parameter's were estimated, was constructed from regions of consecutive ticks (500 regions of consecutive points covering approximately two weeks trading were used). The strength of the diagonal terms demonstrate the meaning reverting. This model was used in the Kalman filter to produce a filtered data set without missing observations, from which subsequent models were produced.

Table 1 show the filtering results for the neural network Kalman filter and the naive random walk hypothesis. The Kalman filter produces superior results for both the mean squared error (MSE) and the robust median absolute deviations (MAD).

Table 1: Model Comparison.

Model	**MSE**	**MAD**
RW Model (No Filter)	2.169 (10^{-4})	0.0084
Neural Network Kalman Filter	1.511 (10^{-4})	0.0075

Figures 3 and 4 demonstrate the Kalman filter identifying outliers. In Figure 3 the filtered estimated states for the $/DM exchange rate are represented by the solid line, the actual observed trades by circles, and trades occurring on the other exchange rates by vertical lines. The filtered states represent the estimate of the true arbitrage value (not the bid or ask values). The effect of market friction's has been incorporated into the estimate $\mathbf{R}_t$, so filtered states are always within the actual bid and ask values observed. The Kalman filter identifies an outlier and uses pure prediction for the estimate of the $/DM rate at time 15.07.41 and does not follow the spurious price movement. The mispricing is filtered and classified as an outlier ($r_t = 4.65$) by the robust algorithm presented here. Figure 3 also demonstrates how new information occurring on any of the exchange rates is immediately incorporated in the robust estimates of all the states. At time 15.07.27, the vertical dotted line indicates the observation of a DM/£ trade, the estimate of the $/£ state is instantly update to incorporate the affect of the rise in the DM/£.

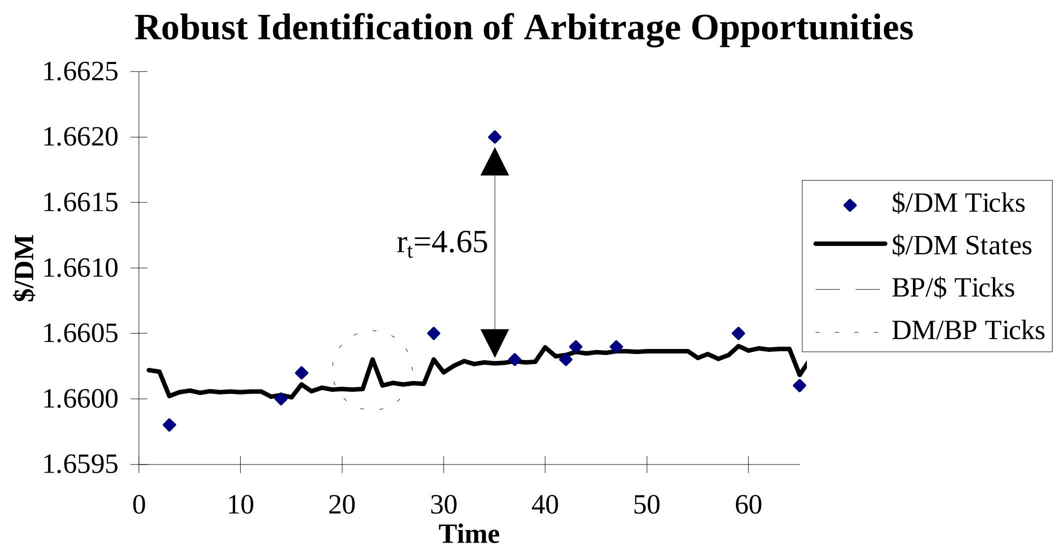

Figure 3: Identification of Market Mispricing.

Table 2: Tick Data (bold) and Estimated Rates (normal).

$/DM Ticks	£/$ Ticks	£/Dm Ticks	$/DM Estimate	£/$ Estimate	£/DM Estimate	SD (10^{-5}) Ln($/D)	SD (10^{-5}) Ln(£/$)
1.6601		**2.4351**	**1.66018**	1.46637	**2.43444**	**4.33**	8.38
			1.66021	1.46624	2.43439	4.21	7.9
1.6604			**1.66031**	1.46627	2.43447	**4.35**	8.18
			1.66030	1.46626	2.43443	4.21	7.71
	1.4662		1.66031	**1.46622**	2.43439	4.65	**8.38**
			1.66031	1.46624	2.43440	4.34	7.7
			1.66031	1.46623	2.43440	4.57	7.9
1.6608			**1.66055**	1.46617	2.43464	**4.6**	8.4
			1.66042	1.46619	2.43449	4.21	8.31
	1.4665	**2.4356**	1.66047	**1.46638**	**2.43488**	4.65	**8.38**

Table 2 demonstrates the ability of the Kalman filter to deal with the erratic arrival of tick data. Bold font exchange rates represent seconds when ticks are observed, and normal font represents the Kalman filter estimates. When incomplete observation vectors are observed the Kalman filter uses the multivariate auto-regressive structure to estimate the unseen rates. In seconds where no observations are observed the filter uses pure prediction to estimate the missing currency rates. The last two columns in table 2 show the estimated prediction standard deviations for the log of the states. Again, bold font indicates the seconds where ticks where observed, and normal font indicates the recursive estimates of the Kalman filter. The prediction error standard deviations grows steadily in periods where no ticks are observed, and collapses to the one second state prediction error standard deviation for the prediction based on new information (ticks).

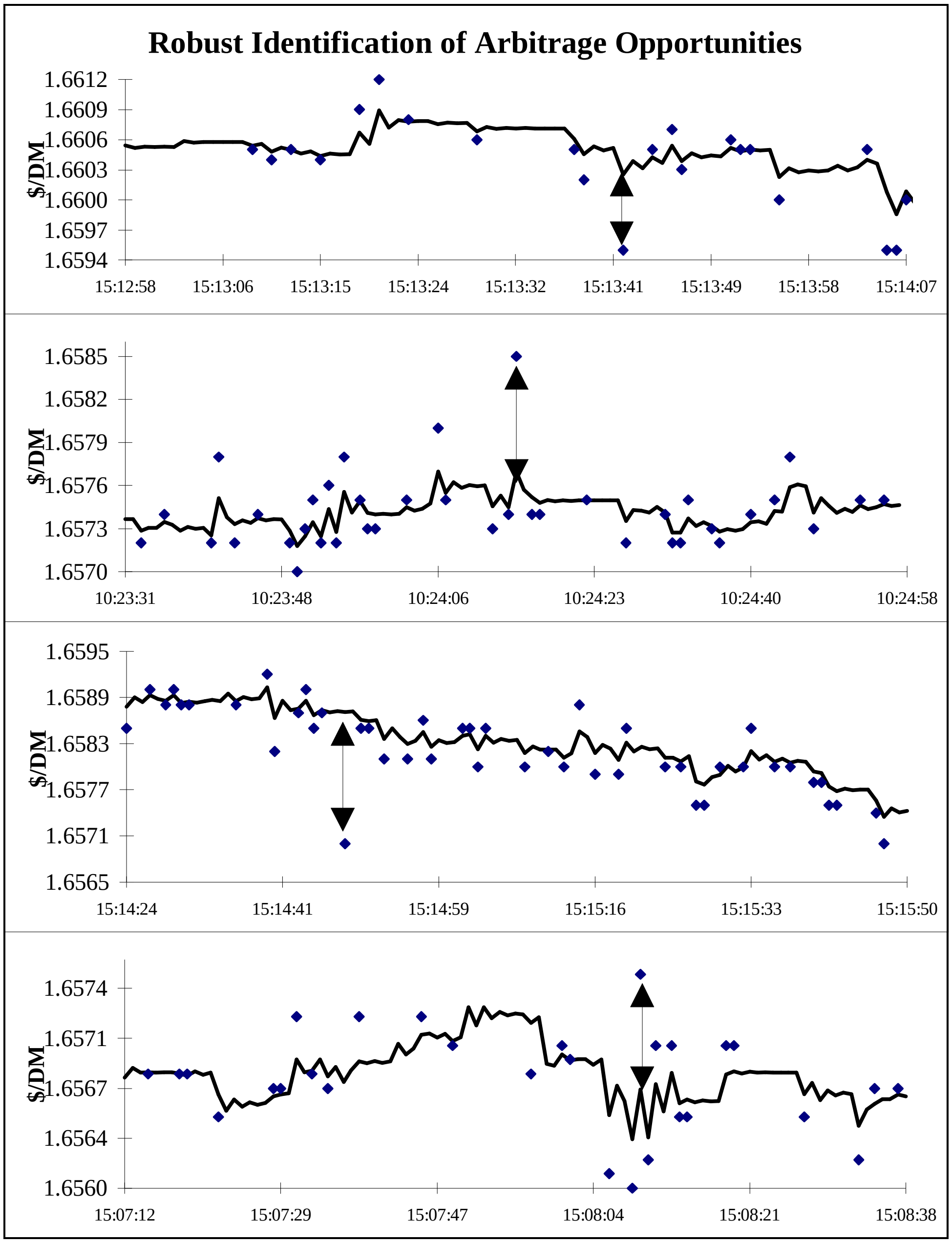

Figure 4: Identification of Market Mispricing.

6. Conclusion

We have shown the effectiveness of one filtering approach for identifying arbitrage opportunities on currency tick data. The methodology is ideally suited to the poor data quality of the financial markets (erratic arrival and additive outliers). The Kalman filter produces rate estimates every second whether or not any ticks are actually observed, this increases both the speed and efficiency of identifying arbitrage opportunities. It is straightforward to extend the above analysis to many more exchange rates and cross rates, increasing the possibility of finding mispricing. In addition this methodology could be readily applied to all forms of arbitrage which are described by similar sets of price relationships.

7. References

Bolland, P. J. and Connor, J. T., "Multivariate Non-Linear Kalman Filters", Technical Report, London Business School, June 1995.

Connor, J. T., Martin, R. D., and Atlas, L. E. “Recurrent Neural Networks and Robust Time Series Prediction”, *IEEE Transactions on Neural Networks*, pp. 240-254, March 1994.

Chung, P. Y., “A Transactions Data Test of Stock Index Futures Market Efficiency and Index Arbitrage Profitability”, *Journal of Finance*, Vol. 46, pp 1791-1809, Dec. 1991.

Dacorogna, M. M., “Price Behaviour and Models for High Frequency Data in Finance”, Proceedings of the NNCM conference, London, England, Oct. 11-13, 1995.

Dempster, A. P., Laird, N. M., and Rubin, D. B. “Maximum likelihood from incomplete data via the EM algorithm.” *Journal of the Royal Statistical Society, B*, 39, 1-38. (1977) .

De Jong, P., “The Maximum Likelihood of a State Space Model”, *Biometrica*, Vol. 75,1, pp 165-169, 1988.

Fama, E. F., “The Behaviour of Stock Market Prices”, *Journal of Business*, Vol. 38, pp 34-105, Jan 1965.

Frenkel, J. A., Levich, R. M. “Covered Interest Arbitrage: Unexploited Profits?”, *Journal of Political Economy*, Vol. 83, pp 325-338, 1975.

Frenkel, J. A., Levich, R. M. “Transaction Costs and Interest Arbitrage: Tranquil Versus Turbulent Periods”, *Journal of Political Economy*, Vol. 85, pp 1209-1226, 1977.

Ghysels, E., Jasiak, J., “Stochastic Volatility and Time Deformation: An Application of Trading Volume and Leverage Effects”, Proceedings of the HFDF-I conference, Zurich, Switzerland, March 29-31, Vol. 1, pp 1-14, 1995,

Harvey, A. C., Ruiz, E., Shephard N. G., “Multivariate Stochastic Variance Models”, Financial Markets Group discussion paper, London School of Economics, 1992.

Huber, P. J. “Robust Statistics”, Wiley, New York, 1981.

Jordan, M.I. and Jacobs, R.A. “Hierarchical mixtures of experts and the EM algorithm”, submitted to *Neural Computation*.

Kalman, R. E. “A new approach to linear filtering and prediction problems.”, *Trans. ASME J. Basic Eng. Series D*, Vol. 82, pp. 35-46, Mar 1961.

Martin, R. D., Samarov, A., Vandaele W. “Robust Methods for Time Series”, In Zellner, A. “Applied Time Series Analysis of Economic Data”, U.S. Bureau of the Census, Washington, pp 153-169, 1983.

Masreliez, C. J. “Approximate Non-Gaussian Filtering with Linear State and Observation Relations”, *IEEE Transactions on Automatic Control*, pp. 107-110, Feb. 1975.

Meditch, J. S. “Stochastic Optimal Linear Estimation and Control”, McGraw-Hill, New York, 1969.

Muller, U. A., Dacorogna, M. M., Oslen, R. B., Pictet, O. V., Schwarz, M., Morgenegg, C., “Statistical Study of Foreign Exchange Rates, Empirical Evidence of a Price Change Scaling Law, and Intraday Analysis”, *Journal of Banking and Finance*, Vol. 14, 1189-1208, 1990.

Rhee, S. G., Chang, R. P. “Intra Day Arbitrage Opportunities in Foreign Exchange and Eurocurrency Markets", *Journal of Finance*, Vol. 47, pp 363-379, Mar. 1992.

Roll, R. "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market", *Journal of Finance*, Vol. 39, pp 1127-1140, Sept. 1984.

Ross, S. A. “The Arbitrage Theory Of Capital Asset Pricing”, *Journal of Economic Theory*, Vol. 13, pp 341-360, Dec. 1976.

Ruiz, E., “Quasi-maximum Likelihood Estimation of Stochastic Volatility Models”, *Journal of Econometrics*, Vol. 63, pp 289-306, 1993.