Does money matter in inflation forecasting?

J.M. Binner [a], P. Tino [b,*], J. Tepper [c], R. Anderson [d,e], B. Jones [f], G. Kendall [g]

[a] Aston University, Birmingham, UK
[b] University of Birmingham, Birmingham, UK
[c] Nottingham Trent University, Nottingham, UK
[d] Federal Reserve Bank of St. Louis, USA
[e] Aston University, UK
[f] State University of New York, Binghamton, USA
[g] University of Nottingham, Nottingham, UK

ARTICLE INFO

ABSTRACT

This paper provides the most fully comprehensive evidence to date on whether or not monetary aggregates are valuable for forecasting US inflation in the early to mid 2000s. We explore a wide range of different definitions of money, including different methods of aggregation and different collections of included monetary assets. In our forecasting experiment we use two nonlinear techniques, namely, recurrent neural networks and kernel recursive least squares regression—techniques that are new to macroeconomics. Recurrent neural networks operate with potentially unbounded input memory, while the kernel regression technique is a finite memory predictor. The two methodologies compete to find the best fitting US inflation forecasting models and are then compared to forecasts from a naïve random walk model. The best models were nonlinear autoregressive models based on kernel methods. Our findings do not provide much support for the usefulness of monetary aggregates in forecasting inflation. Beyond its economic findings, our study is in the tradition of physicists' long-standing interest in the interconnections among statistical mechanics, neural networks, and related nonparametric statistical methods, and suggests potential avenues of extension for such studies.

Crown Copyright © 2010 Published by Elsevier B.V. All rights reserved.

1. Introduction

It is a widely held belief among macroeconomists that there exists a long-run relationship between the growth rate of the money supply and the growth rate of prices (i.e., inflation). This belief forms the foundation for monetary policymaking at the world's central banks, and hence is extraordinarily important for the conduct of public policy. Its importance makes it one of the most commonly tested hypotheses in economics. Yet, the mechanism through which money affects an economy's overall, average price level, is necessarily complex—as complex as the economies themselves. The mechanism almost surely is not linear, and the short-run dynamics may disguise the long-run relationship, confusing tests of the relationship. Linkages need not be univariate, and fluctuations in other variables (including the growth rate of productivity, and international economic conditions) may affect the near- and medium-term correspondence between money growth and inflation. Such interactions raise the possibility that the correspondence may be both nonlinear and time-varying, perhaps with complexity beyond capture in parametric frameworks. In a recent paper, Bachmeier et al. [1], for example, soundly reject both linear autoregressive (univariate) and vector-autoregressive (multivariate) models.

---

* Corresponding address: University of Birmingham, Birmingham B15 2TT, UK.
  E-mail address: P.Tino@cs.bham.ac.uk (P. Tino).

If indeed there exists a dynamic, long-run relationship between the money supply and increases in prices, then it is a reasonable proposition that the near-term growth of the money supply might have predictive power for inflation. This study explores that relationship. In this paper, we investigate a wide range of measures of the money supply (i.e., monetary aggregates) for the United States, and evaluate their usefulness as leading indicators for inflation in the early to mid 2000s. We derive inflation forecasts using two nonlinear techniques with varying degrees of input memory, namely, recurrent neural network operating with potentially unbounded input memory, and a kernel regression technique with finite input memory.

In addition to the economic analysis, our study contributes to physicists' long-standing interest in the interconnections among statistical mechanics, neural networks, and related nonparametric statistical methods. A number of the statistical issues addressed in this paper are similar to those addressed in econophysics. Consider the problem of linear versus nonlinear models. Galluccio et al. [2], Laloux et al. [3], Plerou et al. [4] and Plerou et al. [5], for example, study correlations among returns on different stocks by applying random matrix theory to the stock returns' cross-correlation matrix. This approach imposes a linear model on the statistical linkages among the entities under consideration (e.g., stock returns, monetary aggregates, inflation); the assumed linearity is a limitation of the approach despite the approach being 'model-free' to the degree that the correlation measure is a model-free quantity. Our analysis proceeds in a nonlinear, model-based fashion to consider possible linkages among future inflation, past inflation, interest rates, and monetary aggregates. We assume, specifically, that the structure of such relationships can be explained through two classes of nonlinear models: one with infinite and the other with finite input memory. In addition to their importance in their own right, our findings suggest possibilities for generalizations and extensions of previous econophysics financial-market studies that have used the correlation-based or mutual-information-based method.[1] In this regard, our study adopts the argument made by Plerou et al. [6] that "In economics, one often starts with a model and tests what the data can say about the model. The physics approach to this field differs in that it starts in the spirit of experimental physics where one tries to uncover the empirical laws which one later models." Saito [7], for example, studies the possibility of using a neural network to statistically learn an economic time series—but his network validation was performed using data generated by a model developed in econophysics rather than observed data. Finally, we note that our methodology also lies within the discipline of computational intelligence, a topic of interest to many physicists. Chen et al. [8], for example, illustrate the application of computational intelligence in many fields, including computer science, engineering, psychology and, especially, physics.

2. Monetary aggregates and monetary policymaking

Policymakers worldwide seek to foster economic conditions of robust economic activity and low, stable rates of inflation. In that regard, it is important to identify indicators of macroeconomic conditions that will alert policy makers to impending inflationary pressures sufficiently early to allow the necessary actions to be taken to control and remedy the problem. Equally important, such indicators must not falsely signal future increases in inflation that cause policymakers to slow the pace of economic output in vain efforts to temper inflation. Given the widely held belief in the existence of a long-run relationship between money and prices, monetary aggregates would seem to hold much promise as indicator variables for central banks. The European Central Bank (ECB), for example, employs a "two-pillared" approach, which includes monetary analysis. Specifically, ECB [9] states that "…the [President's Introductory] statement will [after identifying short to medium-term risks to price stability] proceed to monetary analysis to assess medium to long-term trends in inflation in view of the close relationship between money and prices over extended horizons." Evidence to date has not provided strong support for the proposition that monetary aggregates assist forecasting inflation for the United States; see, for example, [10,11].[2] Moreover, as noted by Federal Reserve Board chairman Ben Bernanke, monetary aggregates have not played a central role in the formulation of US monetary policy since 1982. He further states [12]: "Why have monetary aggregates not been more influential in US monetary policymaking, despite the strong theoretical presumption that money growth should be linked to growth in nominal aggregates and to inflation? In practice, the difficulty has been that, in the United States, deregulation, financial innovation, and other factors have led to recurrent instability in the relationships between various monetary aggregates and other nominal variables."

Recently, some economists have issued cautionary notes on the importance of money. See, for example, [13–15]. In particular, Carlstrom and Fuerst [16] state "…we think the current de-emphasis on the role of money may have gone too far. It is important to think seriously about the role of money and how money affects optimal policy." In a similar vein, the Governor of the Bank of England Mervyn King [17] stated "My own belief is that the absence of money in the standard models which economists use will cause problems in future, and that there will be profitable developments from future research into the way in which money affects risk premia and economic behavior more generally. Money, I conjecture, will regain an important place in the conversation of economists."

Linkages between money and inflation have become increasingly important during the recent financial crisis as, in a number of countries, the instrument of monetary policy has shifted towards the quantity of money and away from overnight

---

[1] A generalization to admit nonlinear relationships might be achieved by considering information theoretic measures such as multivariate mutual information.

[2] Stock and Watson [11] also conclude that most "real" economic variables, including real GDP, are not helpful in predicting future inflation.

interest rates (for example, Bank Rate in the UK; the federal funds rate in the United States). A complicating factor in this debate is that money measurement is a complex problem. There are two key questions. First, which monetary assets are to be included in a particular aggregate? And, second, how are these assets to be aggregated together?

Economists generally classify a financial asset as "money" if it is a medium of exchange (that is, may be used for the purchase or sale of goods and services without being converted into any other form), or if the asset may be converted into a medium of exchange at almost any time, and at low cost. This classification scheme admits a wide range of candidate assets. The most obvious is currency. "Narrow" monetary aggregates include currency and bank deposits transferable by cheque. Somewhat broader aggregates include the assets of narrow aggregates but, in addition, assets that are not medium of exchange but have little or no cost of conversion into a medium of exchange; in the United States, these include savings deposits and shares in money market mutual fund. "Broad" monetary aggregates, in addition, include assets with higher conversion costs into a medium of exchange; many of these assets have contractual fixed terms to maturity and financial penalties for withdrawal prior to contract expiration. Among these are small-denomination time deposits. Typically, the assets in narrow aggregates earn zero, or relatively low, interest rates. Broader aggregates, in contrast, include a wide range of assets with higher returns, such as money market mutual funds.

This study considers a number of monetary aggregates for the United States, defined in Appendix. We investigate the forecasting performance of both Divisia and simple sum monetary aggregates at various levels of aggregation ranging from narrow to broad. For comparison purposes, we also explore several interest rate variables. Our forecasting experiments with Divisia monetary aggregates build on several recent papers, including [18–20,1], but are novel in several aspects, discussed below.

We consider a wide range of money aggregates because the choice among the aggregates is far from clear-cut: both the definitions and roles of monetary aggregates in US policymaking have changed through time, largely in response to financial innovation and deregulation. Prior to 1980, both academic researchers and policymakers focused on M1, containing currency and checkable deposits. Following significant deregulation in 1980, attention shifted toward the broader M2 monetary aggregate [21,22]. Yet, the initial stability of its demand (at least, in linear parametric models) proved ephemeral: subsequent studies traced its instability to innovations that changed the behavior of small-denomination time deposits [23]. During both the 1970s and 1980s, the repeated introduction of new substitute financial products caused some analysts to explore the MZM aggregate, consisting of currency plus those financial assets issued by banks and money market mutual funds that could be converted to checkable deposits without cost (that is, all financial assets with "zero maturity"). Duca and Vanhoose [24] state that one response to the instability of M2 demand, at least by researchers if not policymakers, was a renewed focus on MZM, motivated by the assertion that the instability of M2 demand largely was due to heightened substitution between small time deposits and shares in bond-type mutual funds—yet, demand for MZM also encountered instability in the early 2000s. By the mid 1990s, the Federal Reserve largely had abandoned monetary aggregates as an important policy guide. In 1994, the policymaking Federal Open Market Committee began announcing its policy objective to the public after each meeting solely in terms of a desired level of the federal funds rate; in 2001, the Committee ceased publishing annual monitoring ranges for growth of monetary aggregates.

We note that the M1 monetary aggregate used in this study differs from the M1 aggregate published by the Federal Reserve's Board of Governors on its H.6 statistical release. We add to the published M1 the amount of transaction deposits that banks are allowed, under Federal Reserve regulations, to report as saving deposits. The latter reclassification of deposits is referred to as "retail deposit sweeping" [25,26]. Hence, our M1 aggregate corresponds to the M1RS aggregate as defined by Dutkowsky et al. [27]. Data are obtained from the web site of the Federal Reserve Bank of St. Louis (research.stlouisfed.org).

In addition to the uncertainty regarding which financial assets to include within a monetary aggregate, a second measurement issue arises: How should the selected financial assets be aggregated together? There are two candidate methods of aggregation. First, the monetary aggregates (including M1, MZM, M2 and M3) published by the Board of Governors of the Federal Reserve System (Washington, DC) are constructed by simple summation; that is,adding together the amounts of assets as different as currency, saving deposits, time deposits with fixed maturities, and money market mutual fund shares. But a second, alternative method of aggregation is suggested by the economic theory of index numbers. The use of Divisia index number formulae for construction of monetary aggregates was proposed by Barnett [28,29]. Because such monetary aggregates draw on both statistical index number and consumer demand theory, they are theoretically preferable to monetary aggregates built by simple summation. Further, Divisia aggregates admit time variation in liquidity properties of the components of a monetary aggregate by causing the weight of each component within the aggregate to depend on both its expenditure share relative to the other components of the monetary aggregate and on its user cost.[3] (The growth rate of the overall Divisia index is calculated as the weighted average of the growth rates of its components,

---

[3] We note that Divisia indexes are correctly defined in continuous time. The most common discrete time approximation is the Törnqvist–Theil (TT) index [30,31], a chained superlative index [32,33]. Barnett et al. [34] and Anderson et al. [35] provide surveys of the relevant literature, whilst [36] review the construction of Divisia indexes and associated problems.

Divisia economic index numbers provide parameter-free and estimation-free approximations of aggregator functions and are therefore independent of the researcher's choice of theoretical and empirical model. The quality of an index number is dependent on how well it tracks the unknown aggregator function. Since the Divisia index is directly derived from the utility maximization problem it can track the unknown underlying aggregator function exactly. The continuous time Divisia index is directly derived from the rational individuals (current period) blockwise weakly separable sub-utility function over monetary assets. Hulten [37] has shown that the Divisia line integral is path independent under weak separability.

where the weights are each asset's share in the total expenditure on all monetary assets.) Typically, currency and non-interest-bearing deposits receive the highest weight because they are highly liquid assets and (correspondingly) have high user costs (in terms of interest foregone). Interest-bearing time deposits by contrast pay a relatively high rate of interest and, hence, have a smaller expenditure share and a lower weight than might be expected from their relative size. Economic theory suggests that Divisia monetary aggregates should be more closely related to the overall level of liquidity and total spending in the economy than are conventional monetary aggregates.

Despite their grounding in index number theory, measurement issues arise with Divisia monetary aggregate measures. As previously noted, the growth rate of a Divisia monetary aggregate equals the sum of the expenditure share-weighted growth rates of its component assets. Measures of these expenditure shares depends on measures of the opportunity cost of the monetary assets, defined as the spread between the monetary assets' own rates of return and the rate of return on a "benchmark" asset that does not provide any monetary services in the current period (that is, using the asset to purchase goods or services during the current period is infeasible). In practice, the choice of a benchmark rate is uncertain and the choice affects the properties of the aggregate; see, for example, [38]. Drake and Mills [19] suggest that the relatively poor forecasting performance of Divisia M2 in their study may be due to the choice of benchmark rate used by the Federal Reserve Bank of St. Louis in its construction. Hence, a first novel feature of our study is that we investigate how the choice of benchmark rate affects the forecasting performance of Divisia aggregates over a range of different levels of aggregation. We discuss the details surrounding this issue in the next section. A second novel aspect of the paper lies in our use of nonlinear techniques to examine the recent experience of inflation in the US[4] Our previous experience in inflation forecasting using state-of-the-art approaches gives us confidence to believe that significant advances in macroeconomic forecasting and policymaking are possible using techniques such as those employed in this paper. As in our earlier work [41], results achieved using artificial intelligence techniques (i.e., neural networks and kernel methods) are compared with those using a baseline naïve predictor.

3. Data and forecasting framework

Two aspects of the data must be discussed: the choice of an inflation measure and the choice of a monetary aggregate. The choices are substantive because we cannot pretest for invariance (with respect to the measure of money and inflation) of the nonlinear and potentially time-varying relationships that connect a specific monetary aggregate and an inflation measure.

3.1. Choice of inflation measure

A number of inflation measures are available for the United States, including the chain-price index for gross domestic product, the chain-price index for personal consumption expenditures both including and omitting food and energy, the implicit price deflator for gross domestic product, and several variants of the consumer price index for urban workers and wage earners (CPI-U and CPI-W). The CPI-U is perhaps the best known among households and businesses. Although the Federal Reserve's policymaking Federal Open Market Committee monitors the chain-price index for personal consumption expenditures omitting food and energy ("core PCE price index") on the belief that it best measures the underlying inflation trend affected by monetary policy, the Committee's underlying concern, as summarized by a former Federal Reserve governor, is the path of headline CPI inflation: "I have no qualm in stating that controlling headline inflation, not core inflation, is—along with maintaining maximum sustainable employment—the ultimate aim of monetary policy." [42] Similarly, inflation-targeting central banks worldwide focus on the behavior of headline CPI measures, regardless of their intermediate price index targets [43]. Annual cost-of-living adjustments for Social Security recipients, as well as federal government and military retirees, are based on the CPI-U, as are annual adjustments of federal income tax brackets and the payments to holders of the Treasury's inflation-protected (indexed) bonds.

In its construction, the CPI-U is a compromise between rapid evolution of expenditure weights so as to accurately track changing consumer expenditure patterns and the more sluggish adjustment required to measure cost-of-living changes relative to a set of base-period reference expenditure shares. In the event, CPI-U expenditure weights are updated every two years, based on consumer surveys. At finer levels of disaggregation, goods and services are aggregated using geometric means so as to minimize substitution bias; at more aggregate levels, aggregation is via Laspeyres formulas so as to avoid understating the impact of rising prices on households' cost of living. The CPI has a careful process for quality adjustment, partially based on hedonics that seeks to include inevitable changes in available goods and services. Housing is treated via a rental equivalence approach that both collects actual spending through the Consumer Expenditure Survey and market data regarding the rental prices of apartments and houses. Greenleens and McClelland [44] provide an accessible discussion of these and other issues.

An alternative index used in some studies, and also available monthly, is the chain-price index for PCE. This index is closely related to the CPI in that both indexes are built from the same underlying price and quantity data furnished by the US Bureau of the Census, and both in their current versions measure the user cost of housing services consumed by households rather

---

[4] A number of recent studies of inflation and inflation uncertainty have appeared in the literature, including [39,40].

than the purchase prices of houses, utility services, etc. Yet, there are substantive differences between the indexes, including the formula used for upper-level aggregation (the CPI uses a Laspeyres formula, the PCE chain-price index uses a Fisher-Ideal superlative formula), the weights (the CPI uses weights from a monthly consumer expenditure survey, the PCE index uses weights from monthly business surveys including the survey of retail sales), the updating of the weights (CPI weights are updated every two years, PCE index weights are updated each year based on a ratio of the preceding two years), quality adjustments (mandated changes due to safety or environmental concerns generally appear as price increases in the CPI but as quality adjustments in the PCE index), and scope (the CPI includes only out-of-pocket payments by consumers, while the PCE includes payments made by firms on behalf of employees). For the period considered in this study (1960–2005), the two indexes behave similarly: average monthly year-over-year inflation measured by the CPI-U and PCE index is 4.3% and 3.8%, respectively, and the differences between the measure in almost all months is less than 1% except for periods in 1979–1981 when CPI inflation exceeded 10% due to elevated mortgage interest rates. In empirical studies, use of the PCE price index is not obviously to be preferred to the CPI. Clark [45] argues that the CPI is preferable because the data are more timely and of better quality. Although his analysis is a decade old, it seems the situation has changed little since [46,47].

Price indexes at lower frequencies also are available, but we do not consider them in this study.[5] The chain-price index for GDP is available quarterly, but with a longer lag than the CPI; the first observation for a quarter is published late in the third month following the end of the measured quarter, revised values are published late in each of the following two months, and a further value is published the following year. Its delayed availability and tendency to be revised make the GDP price index less informative to policymakers and it is infrequently used in contracts. The GDP price index both contains items excluded from the CPI and includes additional items. Excluded items include prices of used consumer goods and prices of imported goods. Many of the included items are intermediate products, including consumer goods in distribution and producers' durable equipment and structures. Changes in such prices are only weakly linked to changes in consumer goods' prices and the cost of living.[6]

Inflation in this paper is measured monthly as the year-over-year percentage change in the CPI-U. We prefer the CPI due to its wide-spread public recognition and timely data availability. In the interests of brevity, we do not explore alternative measures here but leave them as a subject of future research.

3.2. Monetary aggregates

We consider in this study 16 alternative monetary aggregates, varying in terms of the monetary assets included in each aggregate and the method of aggregation.

First, we consider four different groupings of monetary assets: M1, which consists of currency, demand deposits, and other checkable deposits; MZM, which consists of all M1 assets plus savings deposits, including money market deposit accounts (MMDAs) and retail-type money market mutual funds; M2, which consists of MZM assets plus small-denomination time deposits; and M3, which consists of M2 assets plus institutional-type money market mutual funds, repurchase agreements, Eurodollar accounts, and large-denomination time deposits.

Second, we consider monetary aggregates built with different aggregation methods. We obtained simple sum and Divisia monetary aggregates for each of the four levels of aggregation (M1, MZM, M2, and M3) from online databases maintained by the Federal Reserve Bank of St. Louis; see [50]. In addition, we obtained the Divisia monetary aggregates built by Elger et al. [20] at the same four levels of aggregation (M1, MZM, M2, and M3). These aggregates differ from those constructed in St. Louis primarily in two respects: the choice of the benchmark rate that is used in measurement of assets' opportunity costs and the imputed own rate of return on non-interest-bearing bank deposits. Elger et al. [20] choose the benchmark rate as the upper envelope of the returns in M3; that is, in each time period the benchmark rate is measured as the maximum own-interest rate on the assets included in the M3 level of aggregation; the St. Louis aggregates are based on an upper envelope that includes a long-term bond rate (the Baa rate) that, in practice, is selected as the benchmark rate in the majority of time periods. In addition, Elger et al. [20] set to zero the own-rate of return on non-interest-bearing demand deposits, while the St. Louis indexes include an implicit return. (There are additional differences between the aggregates not discussed here; see [20, pp. 432–3].) Finally, as a check for robustness, we construct an additional set of Divisia aggregates that follow the construction method of [20] except that the Baa corporate bond rate is included in the upper envelope selection process for the benchmark rate.

Finally, we also considered in our study the inflation forecasting potential of both a short-term and a long-term interest rate: specifically, the bond-equivalent yield on three-month Treasury bills and the yield to maturity on Baa corporate bonds. We allowed each model the freedom, if it wished, to include these rates alongside or instead of the chosen measures of money, or to omit the interest rates altogether. Each model was permitted to select lags and/or orders of differencing of each variable as it wished. By permitting the model to choose among such transformations, we also implicitly include the slope of a yield curve in our model, a variable that some previous studies have concluded has predictive power for inflation [51,52]. In no case did the inclusion of interest rates improve a model's performance.

---

[5] Nakamura [48], for example, studies the quarterly GDP implicit deflator.

[6] Nakamura [49] concludes that variation in retail prices arises primarily from retail-level demand and supply shocks rather than manufacturer-level price fluctuations.

4. Methodologies

In our experiments, two artificial intelligence forecasting techniques—recurrent neural networks and kernel recursive least squares—compete to predict the year-over-year inflation rate that will prevail six months in the future. For each technique, we begin with a competition in which a large number of models are fit to the training dataset; we retain the four best models for each technique judged on their performance on the hold-out validation dataset. Next, we compare the forecasting performance of the four models (and their flat ensembles) for out-of-sample forecasting on the test set. Of the 541 data points available, the first 433 were used for training (model fitting) (January 1960–February 1997), the next 50 data points were used for validation (March 1997–April 2001), and the next 46 data points were used for forecasting (May 2001–February 2005). We also consider a "naïve" random walk model for comparison. Neural network and kernel techniques are able to approximate a wide family of linear and nonlinear data transformations, and hence are robust to the order of integration of explanatory variables [53–55]. The input data series, i.e., the right-hand side explanatory variables for our experiments, in the jargon of time-series econometrics, are allowed to be of differing "orders of integration." This is addressed further below.

The ability of artificial intelligence techniques to approximate linear and nonlinear transformations of the data may be illustrated by considering the least squares regression problem. Consider choosing a vector of fixed parameters, β, so as to solve $\min_\beta \sum_{i=1}^{t}(y_t - \hat{y}_t)^2$, where y denotes the dependent variable; $\boldsymbol{x} = (x_1, \ldots, x_K)$ denotes a vector of explanatory variables $k = 1, \ldots, K$; $\hat{y} = \beta \cdot \phi(x)$; and $\phi(x) = (\phi_1(x), \ldots, \phi_J(x))'$ is a fixed, finite-dimensional mapping defined over the explanatory variables. Linear least squares regression defines $\phi_j(x) = x_k$ for all $\forall j, k$. Regressions with differences of the explanatory variables are denoted $\phi_j(x) = \Delta x_k$; regressions with interaction effects are denoted $\phi_j(x) = x_{k_1} \cdot x_{k_2}, k_1 \neq k_2$, for some j; and regression-based specification error tests are denoted $\phi_j(x) = x_k^2$ or $\phi_j(x) = x_k^3$ for some j (typically, $j > K$). Our techniques admit all of these and others. Kernel regression methods, for example, may be written as $\hat{y}_t \equiv \sum_{j=1}^{J} \beta_j k(x_j, \boldsymbol{x})$, where $\{x_j\}_{j=1}^{J}$ are the training data points and the $k(\cdot, \cdot)$ are positive-definite (nonlinear) kernel functions; see [56] or [57]. Some authors, e.g., Andrews [58], have labelled these nonparametric kernel estimators of semi-parametric models.

An earlier investigation is that in [59]: "We use an artificial neural network [ANN] to explore the explanatory power of a variety of measures of a single variable that is certainly related to inflation and money…the ANN can approximate any Borel measurable function to any degree of desired accuracy …in other words, the ANN can provide a completely flexible mapping that can approximate highly nonlinear functions to any degree of desired accuracy as long as a sufficient number of hidden nodes are included." Dorsey considers inflation to be an unknown function of k recent levels of the money stock, $\pi_t = F(M_{t-1}, M_{t-2}, \ldots, M_{t-k}) + \xi_t$. The function $F(\ )$ is unknown. His methods are capable of suggesting that the preferred $F(\ )$ is linear in levels

$$\pi_t = F(\ ) = \sum_{i=1}^{k} \alpha_i M_{t-i} + \xi_t,$$

or in first differences

$$\pi_t = F(\ ) = \sum_{i=1}^{k} \beta_i \Delta M_{t-i} + \xi_t$$

(where $\Delta M_{t-i} = M_{t-i} - M_{t-i-1}$), or linear in a combination of levels and differences, or nonlinear. Note, in particular, that the flexibility of neural network and kernel methods obviates the need to determine in advance, for example, the order of integration of the explanatory variables. As in [59], when provided with the levels of the "explanatory" variables, these techniques are sufficiently powerful to explore transformations of those variables. Our methods are robust to issues such as the order of integration of the variables and the risk of accepting "spurious regressions".

The power of neural network approximations, for example, is summarized in the well-known theorem of [60] that says, essentially, that a neural network can approximate all probability density functions supported compactly on the real numbers. This version is from [61, p. 245–6]:

Theorem. *Let G be a continuous function, $G : R \to R$, such that either (1) $\int_{-\infty}^{\infty} G(x)\mathrm{d}x$ is finite and nonzero and G is L^p for $1 \leq p < \infty$ or (2) G is a squashing function $G : R \to [0, 1]$; that is, G is nondecreasing, $\lim_{x \to \infty} G(x) = 1$, and $\lim_{x \to \infty} G(x) = 0$. Define the set of all possible single hidden-layer feed-forward neural networks, using G as the hidden layer activation function:*

$$\sum^{n}(G) = \left\{ g : R^n \to R | g(x) = \sum_{j=1}^{m} \beta_j G(w^j \cdot x + b_j); b_j, \beta_j \in R; w^j \in R^n, w^j \neq 0; m = 1, 2, \ldots \right\}.$$

Then, for all $\varepsilon > 0$, probability measures μ, and compact sets $K \subset R^n$, there is a $g \in \Sigma^n(G)$ such that $\sup_{x \in K} |f(x) - g(x)| \leq \varepsilon$ and $\int_K |f(x) - g(x)|\mathrm{d}\mu \leq \varepsilon$.

4.1. Multi-recurrent neural networks

Recurrent neural networks (RNNs) are typically adaptations of the traditional feed-forward multi-layered perceptron (FF-MLP) trained with gradient-descent learning algorithms (see [62,63]). This class of RNNs extends the FF-MLP architecture to include recurrent connections that allow network activations to feed back as inputs to units within the same or preceding layer(s). Such internal memories enable the RNNs to construct dynamic internal representations of temporal order and dependencies which may exist within the data. Units that receive feedback values are often referred to as *context* or *state* units. Also, assuming that nonlinear activation functions are used, the universal function approximation properties of FF-MLPs naturally extends to RNNs. These properties have led to wide appeal of RNNs for modeling time-series data, [64–68].[7]

We build upon our previous successes with RNNs [67,70] for modeling financial time-series data and extend our RNN models to a subset of RNNs referred to as *multi-recurrent networks* (MRNs) [66,71]. The MRN architecture combines several types of feedback and delay to form a state-based model whose state transitions are modeled as an extended nonlinear autoregressive moving average (ARMA) process [71]. The architecture employs three levels of feedback, allowing recurrent connections from (i) the output layer back to the input layer, as found in *Jordan networks* [71]; (ii) the hidden layer back to the input layer, as found in simple recurrent networks *(SRNs)* [72] and, finally, (iii) from the context units within the input layer back to themselves (self-recurrent links).

As described in Section 2, the external input vector consists of between one to four input variables representing combinations of previous inflation rates, one of the measures of money and zero, one or two of the interest rate measures. We do not permit recurrent or self-recurrent connections from external input units. We extend each allowable feedback by additional banks of context units (memory banks) on the input layer. The number of additional memory banks relates directly to the degree of granularity at which past and current information is integrated and stored. Following [66], we use φ memory banks where $\varphi = 4$, as past experience has shown that moving beyond this does not lead to enhanced performance. Rather, it is the number of units within each bank that is pivotal to the performance of the network and that is determined on the validation set. The MRN architecture is shown in Fig. 1, and it can be expressed generally as

$$\hat{y}(t + \tau) = g(f(\mathbf{c}(t), \mathbf{x}(t), \mathbf{W}_f(t)), \mathbf{W}_g(t)), \tag{1}$$

where $\hat{y}(t + \tau)$ denotes the predicted value of inflation (the target for is $y(t + \tau)$ where $\tau = 6$ for our purposes and refers to the value of inflation six months ahead); $\mathbf{c}(t)$ (the context vector) is the concatenation of the previous hidden state vector with four delays of varying strength and summation of elements of the previous output vector with four delays of varying strength; $\mathbf{x}(t)$ is the external vector of input variables; $\mathbf{W}_f(t)$ is the weight matrix connecting the input layer to the hidden layer, $\mathbf{W}_g(t)$ is the weight matrix connecting the hidden layer to the output layer and, finally, vector function f and function g return the activation vectors from the hidden and output layers, respectively. We apply the hyperbolic tangent function to the inner products performed for f and the binary sigmoid function to the inner products performed for g. We also use an *internal* output unit to provide an estimate, $\hat{\varepsilon}(t + \tau)$, for the noise term $\varepsilon(t + \tau)$, where $\hat{\varepsilon}(t + \tau) = y(t + \tau) - \hat{y}(t + \tau)$. For the inner products associated with $\hat{\varepsilon}(t + \tau)$, we use the identity function for g.

After a number of preliminary experiments to determine the most useful information to feed back from the output layer, it was clearly evident that feeding back the target value, $y(t + \tau)$, produces superior models (with respect to performance on the validation set) compared to feeding back either the predicted forecast, $\hat{y}(t + \tau)$, or the estimated noise value, $\hat{\varepsilon}(t + \tau)$. This method, however, is unrealistic as, for the current prediction task, the target values would not be available for use during estimation. It was therefore decided to assess whether we could estimate the target, for feedback purposes, by simply summing the predicted forecast with the estimated noise value to form an *internal target*, $\hat{y}'(t + \tau)$. This consistently produced superior models (with respect to performance on the validation set) compared to those which fed back either the predicted forecast or the estimated noise term, and it was therefore selected as the information to feed back from the output layer. As shown in Fig. 1, we return $\hat{y}'(t + \tau)$ as the network's prediction value.

The context vector, $\mathbf{c}(t)$, represents an internal memory of varying rigidity—that is, some context units will represent information from very recent time steps and thus change rapidly whilst others will represent information further back in time and change much more slowly. To achieve this, the unit values fed back from the hidden and output layers are combined, to varying degrees, with their respective context units, $\mathbf{c}_i(t)$, in the input layer (as determined by the weighting values applied to their recurrent links). The influence of the previous context unit values are determined by the weighting values applied to the self-recurrent links. When the weightings on the recurrent links are greater than those on the self-recurrent links then more flexible memories are formed storing more recent information at the expense of historical information stored in the context units. Conversely, if the weighting values on the self-recurrent links are greater than those on the recurrent links then more rigid memories are formed since more historical information is preserved in the context units at the expense of the more recent information being fed back from subsequent layers. This can be generally expressed as follows:

$$c_i(t) = f(a_j(t - 1)v_j + c_i(t - 1)z_i), \tag{2}$$

---

[7] Previous studies have found little forecasting improvement from feed-forward networks. Stock and Watson [69] hypothesize that recurrent networks might do better.
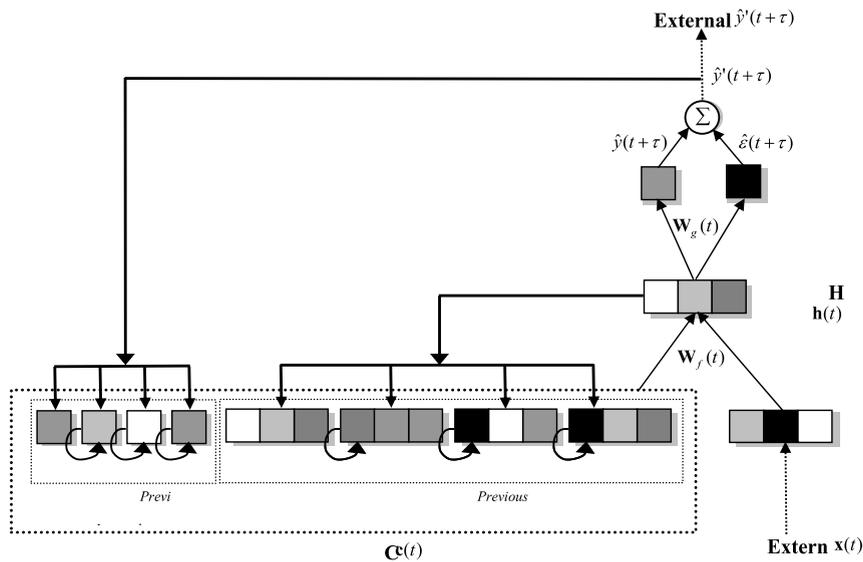
External $\hat{y}'(t+\tau)$

$\hat{y}'(t+\tau)$

$\hat{y}(t+\tau)$ $\qquad$ $\hat{\varepsilon}(t+\tau)$

$\mathbf{W}_g(t)$

H
$\mathbf{h}(t)$

$\mathbf{W}_f(t)$

Previ $\qquad$ *Previous*

$\mathbf{C}(t)$ $\qquad$ **Extern** $\mathbf{x}(t)$

Fig. 1. Architecture for the multi-recurrent network.

where $c_i(t)$ refers to context unit i at time t; $a_j(t-1)$ denotes either an internal output target value or a hidden unit activation value at time $t-1$; v_j refers to the connection strength of the recurrent link from a_j to c_i, where $v_j = \frac{1}{\varphi}j$, where $j = 1, 2, \ldots, \varphi$; $c_i(t-1)$ refers to context unit i at time $t-1$; and, finally, z_i refers to the connection strength of the self-recurrent link for c_i, where $z_i = \frac{1}{\varphi}i$, where $i = 1, 2, \ldots, \varphi$. The number of hidden units, and thus size of the recurrent layer, is a significant hyper-parameter for this model and is determined by performance on the validation set.

We use back-propagation through time [73,74], an efficient gradient-descent learning algorithm for training RNNs. Input data were scaled to zero mean and unit variance. To facilitate convergent learning, the '*search and converge*' learning rate schedule as defined by Darken et al. [75] was used for all training experiments with an initial learning rate, η_0, of 0.0054. This provides an effective time-varying learning rate that guarantees the convergence of stochastic approximation algorithms and has proven effective for temporal domains [76]. The momentum term is fixed at 0.95 due to the low learning rates. To identify the onset of over-fitting during training we use the decline in performance on the hold-out validation set as the basis for terminating the training process. We generate independent training sequences directly from the time series using a time window, whose lag size is determined by performance on the validation set. The MRN context units are initialized to known values at the beginning of each sequence of data, and each element within a sequence is processed sequentially by the network. Although this restricts the MRN to behaving as a finite memory model and also means that the MRN must first learn to ignore the first context values of each sequence, it allows for efficient randomized learning.

Although such RNNs and variants thereof have enjoyed some success for time-series problem domains we tread with cautious optimism. Due to the additional recurrency, such models inherently contain large degrees of freedom (weights) which may require many training cases to constrain. Also, gradient descent-based learning algorithms are notoriously difficult at finding the optimum solution the architecture is theoretically capable of representing; see [77,78].

4.2. Kernel-based regression techniques

Kernel methods (sometimes known as "kernel machines") are a popular class of learning algorithms based on a simple idea: it is possible to produce nonlinear versions of traditional linear learning techniques by formulating them in a (so-called) feature space H (as opposed to working directly in the input space X where the inputs come from). Given a kernel function $K : X \times X \to R$ (satisfying Mercer conditions), there exists a Hilbert space H where K acts as a dot product. In kernel machines, the feature space associated with kernel K is such a Hilbert space H. Now, because many linear learning algorithms can be written purely in terms of dot products of inputs, it is easy to formulate their nonlinear versions using the "kernel trick": simply substitute dot products of inputs by dot products of images of inputs in H, which can be calculated directly using the kernel function K. Kernel machines were shown in numerous studies to be able to achieve competitive state-of-the-art performances. There is one prominent problem with kernel machines, though: they typically scale rather unfavorably with the number of training examples, as this is directly related to the number of degrees of freedom. Numerous approaches for achieving sparsity of kernel machines have been suggested. For more information on kernel machines we refer the interested reader to [57].

Kernel machines have shown great promise in financial forecasting [79–81]. In this paper, we use the "kernelized" version of the linear recursive least squares technique, termed *kernel recursive least squares* (KRLS) [57]. This particular approach is an example of a kernel machine with online constructive sparsification methodology. Loosely speaking, a training input

is allowed to enter construction of the kernel machine only if its image in the feature space H cannot be "approximately" represented as a linear combination of images in H of previously admitted training inputs. This makes sense, since there is no nonlinearity in the model formulation once we transform the inputs into the feature space H. Of course, the resulting model will be nonlinear, because the mapping of inputs to their images in H is nonlinear. The amount of tolerance in allowing the image in H of the currently considered input not to lie exactly in the linear subspace spanned by images in H of the previously admitted training inputs is quantified by a parameter $\nu > 0$. The greater the value of parameter, ν, the more sparse will be the resulting model. On top of this methodology, a kernel-based version of the recursive least squares algorithm operates, enabling it to solve efficiently in a recursive manner nonlinear least squares prediction tasks.

The input vector $x(t) = (x_1, x_2, \ldots, x_L)$ at time step t consists of lagged inputs for the most recent previous L periods, and $x_j, j = 1, 2, \ldots, L$, denotes the value of the input variable at time $t - j + 1$. The resulting (trained) model returns for each input $\mathbf{x}(t)$ a real-valued output $f(\mathbf{x}(t)) \in R$ (predicted inflation rate at time $t + \tau$) that takes the typical form of a kernel regression machine:

$$\hat{y}(t + \tau) = f(\mathbf{x}(t)) = \sum_{i=1}^{N} w_i K(\{\mathbf{x}^j\}, \mathbf{x}(t)) \tag{3}$$

where N is the number of admitted training inputs $\{x^j\}, j = 1, 2, \ldots, N$, and $K(\ldots)$ is the kernel function. Given the set of training input–output pairs $\{x^j, y^j\}, j = 1, 2, \ldots, M$, and due to the model sparsity, N is typically (much) smaller than the number of training inputs M. We used spherical Gaussian kernels, defined as

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}'\|^2}{2\kappa^2}\right), \tag{4}$$

where $\kappa > 0$ determines the kernel width (KW) and $\|\ \|^2$ is the squared Euclidean norm in the input space X. Small kernel widths κ yield more "flexible" models (that are more prone to possible over-fitting of training data).

Given the training data, the collinearity threshold, ν, and the kernel width, the KRLS returns a (potentially) sparse model that can be tested on the hold-out validation set. The most appropriate values for ν and κ are given by the best performing model on the validation set.

We also considered kernel partial least squares (KPLS) [82]. This is a kernelized version of the partial least squares technique. Again, spherical Gaussian kernels were used. The kernel width and the number of latent factors were set on the validation set.

4.3. Naïve model—Random walk hypothesis

Our baseline naïve predictor operates as follows: if the prediction horizon is τ months (in our case $\tau = 6$), predict that in τ months we will observe the current inflation rate $y(t)$. In other words, we predict that the inflation rate at time $t + \tau$ will be $y(t)$, i.e., $y(t + \tau) = y(t)$. This model corresponds to the random walk (RW) hypothesis with moves governed by a symmetrical zero-mean density function, and thus measures "the degree to which the efficient market hypothesis applies."

5. Results

Before reporting the results of individual prediction techniques, we briefly describe performance measures that form the basis of our model evaluation and comparison.

5.1. Performance measures

The actual and predicted inflation rates at time t are denoted $y(t)$ and $\hat{y}(t)$, respectively. Given a sample of T test times $t_1, t_2, \ldots, t_T$, we calculate the root mean square error (RMSE) of the model predictions

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{n=1}^{T} (y(t_n) - \hat{y}(t_n))^2}. \tag{5}$$

However, it may be difficult to appreciate the value of RMSE for model M on its own. Therefore, we also report the rate (in per cent) of improvement in RMSE of the tested model M with respect to the baseline random walk (RW) assumption:

$$\text{IORW(M)} = \frac{\text{RMSE(RW)} - \text{RMSE(M)}}{\text{RMSE(M)}} \tag{6}$$

where RMSE(M) and RMSE(RW) are the RMSE of the model M and RW baseline, respectively. Negative values of IORW(M) indicate the case where model predictions are worse than baseline prediction capabilities.

Table 1
The four best MRN models and the input variables used by each model.

MRN	Money measure	Treasury bill interest rate	Baa interest rate	Past inflation rate
1	Divisia M1–Baa envelope	Yes	No	Yes
2	No	No	Yes	Yes
3	Simple Sum M3	Yes	No	Yes
4	No	Yes	Yes	Yes

Table 2
Number of hidden layers, parameters and RMSE on the test set for the four best RNN models.

MRN	Number of hidden layers	Number of parameters	RMSE
1	25	2,752	0.0072
2	5	147	0.0077
3	10	502	0.0070
4	10	502	0.0064
Ensemble	*n/a*	*n/a*	0.0065

Table 3
The four best kernel models and the input variables used by each model.

KRLS	Money measure	Treasury bill interest rate	Baa interest rate	Lagged inflation rate
1	No	No	No	Yes
2	No	No	No	Yes
3	No	No	No	Yes
4	No	No	No	Yes

5.2. Multi-recurrent neural networks

To determine the combination of allowable input variables which have the most influence on predicted inflation values, a simple selection scheme was followed. First, experiments with only past inflation rates were performed. Past inflation rates are naturally used in all other input variable combinations. Second, experiments with one or both of the interest rate measures (Treasury bill (TB) and Baa) were performed. Finally, to determine the influence of money measures, a measure of money was included as an input variable with zero, one or both of the interest rates. For each of the above variable combinations, time lags of 6, 12, 18, 24 and 32 months were examined. The MRN architecture and training regime described in Section 3 were used in all experiments. The number of allowable free parameters (and thus hidden nodes) was limited to a percentage of the number of training patterns.

After training the MRNs with the above variable combinations, the four best models obtained were as shown in Table 1, and they were all achieved with a time lag of 24 months which generated a training set of 9,864 patterns (411 sequences). MRN models were selected based on performance on the hold-out validation set. It can be seen that in two of the best performing models, models 1 and 3, measures of money and an interest rate did appear to influence the predicted inflation values. However, the level of influence is unclear, as superior performance is evident without a measure of money, namely for model 4, which only used both interest rates and past inflation rates. After repeating experiments with the model 4 configuration, it was clear that over the six-month forecast horizon, superior performance is consistently obtained when using no measure of money, both interest rates and past inflation rates.

Table 2 shows the number of free parameters and resulting performance on the test set for each winning model. Model 1 required substantially more free parameters than other "winning" models. Although this model performed best on the validation set, it performed worst on the test set. Generally, however, very good performance can be obtained on the test set with MRN models that have as low as 5% of the number of free parameters of model 1. We also averaged the predicted inflation values across all four MRN models to obtain the ensemble MRN performance. The resulting RMSE for the validation set proved superior to that of each of the individual MRN models. An illustration of model 4 MRN predictions over the test period is presented in Fig. 2. The solid and dashed lines correspond to the real and predicted inflation rates, respectively.

5.3. Kernel machines

In general, over the six-month forecast horizon, the performances of KRLS and KPLS were comparable; for brevity, only results for KRLS are reported. The best performing KRLS model candidates on the validation set were consistently (nonlinear) autoregressive models on inflation rates alone (see Table 3). Input data was normalized to zero mean and unit standard deviation (the mean and variance are determined on the training set alone). As Table 4 shows, the best four candidates from the KRLS model class used lags of 10 and 12 months. The kernel width κ varied between 1.2 and 1.5 and the collinearity threshold ν ranged between 0.21 and 0.28.

The rather limited range of values of structural parameters for the KRLS class indicates that a stable regime exists within which the KRLS class yields a satisfying performance on the given problem. It seems that having historical monthly inflation
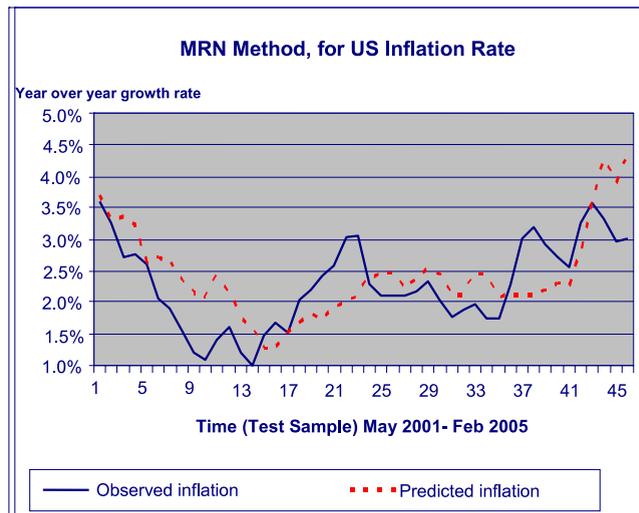
Fig. 2. The best predicted inflation rates obtained by the MRN.

Table 4
Model structure and test set performance of the best four kernel recursive least squares models.

KRLS	Inflation rate lag	KW	ν	RMSE
1	12	1.2	0.27	0.0050
2	10	1.2	0.27	0.0053
3	10	1.2	0.28	0.0057
4	10	1.5	0.21	0.0082
Ensble	n/a	n/a	n/a	0.0055

Table 5
Results summary.

		MRN	KRLS
	Model	4	1
	RMSE	0.0064	0.0050
	IORW	27.5	43.4
	Money?	No	No

rates roughly a year into the past and a spherical kernel of width slightly larger than the standard deviation of the input data, with modest model sparsity induced by $\nu \cong 0.27$, yields the best performance. An illustration of KRLS predictions over the test period is presented in Fig. 3. The solid and dashed lines correspond to the real and predicted inflation rates, respectively.

5.4. Model comparison

A summary of results showing the best fitting model for each approach is reported in Table 5. The kernel-based regression provides the biggest improvement (43.4%) over the random walk, with the recurrent neural network in second place.

Our early indications suggest that the kernel method appears to implement a more efficient feature extraction method that is well suited to the problem. It appears that using the "kernelized" version of the linear least squares technique provides an effective mechanism for constructing efficient feature domains well suited to the task of prediction. One key difference between the two techniques employed in this paper is the treatment of the input variables. The kernel-based regression methods treat the inputs as inherently static, whilst the recurrent neural networks treat the inputs as dynamic due to the internal state memory being formed and fed back at each time step.

Since the validation and test sets are only samples, the performance of the best model on the test set may overestimate the typical performance one would expect from the corresponding model class. The average performance of the few best performing models, however, gives an indication of the expected performance, if sound model selection techniques are used. The average IORW measures of the best four models were 19.9% and 29.7% for the MRN and KRLS model classes, respectively. This gives a strong positive indication of the predictive capacity of the model classes considered in this inflation prediction task.
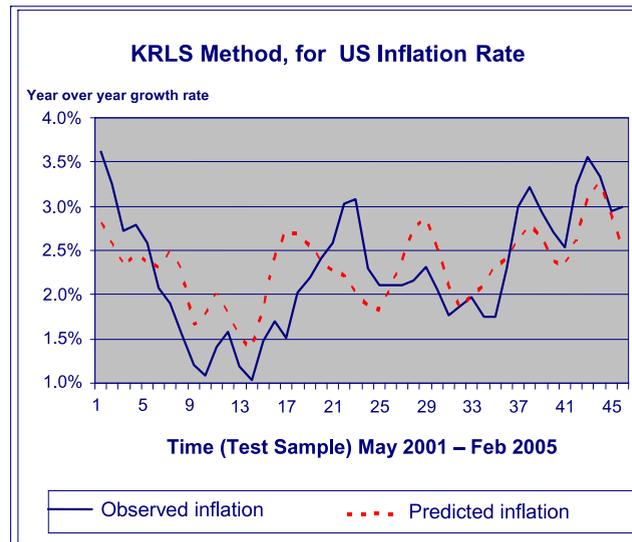
Fig. 3. The best predicted inflation rates obtained by the KRLS.

With respect to the question 'does money matter?', simple sum money and a Divisia money constructed using the upper envelope approach are among the best performing models in the MRN model class. Nevertheless, the best model in that class did not contain any money but did contain interest rates. Overall, the best models are the nonlinear autoregressive models based on kernel methods. It is felt that further improvements in the construction of the money supply might help to obtain improvements in the inflation forecasting performance of the monetary aggregates.

We also considered all the monetary aggregates together as 16-dimensional input time series based on which the future inflation rate was predicted. To base the prediction on the dominant factors (in the variance maximization sense) in the 16 monetary aggregates, we decomposed the 16-dimensional input series using principal component analysis (PCA).[8] We discovered that when using PCA to reduce the signal dimensionality, similar large-scale trends in the monetary aggregates tend to mask the time-conditional differences between them that, however, may be crucial for inflation prediction. There was only a single dominant eigenvalue of the monetary aggregates' covariance matrix, corresponding to the eigenvector of roughly equal terms across all 16 monetary aggregates. Projection onto this eigenvector roughly corresponds to computing the mean of the 16 monetary aggregates. Hence, PCA performed smoothing of the monetary aggregates into a single average aggregate. Predictions based on such an average aggregate were comparable to (or worse than) predictions based on the individual monetary aggregates. This approach, commonly termed principal component regression (PCR), is known to be suboptimal when only concentrating on dominant eigendirections [83]. We tried various combinations of the major component with the minor ones (or minor components only), but the performance did not improve.[9]

6. Discussion, conclusions and future work

Our findings provide the most comprehensive evidence to date that neither the aggregation method nor level of aggregation has a significant impact on the performance of our model, in contrast to the results reported by Schunk [18]. To the contrary, our results lend more support to the conclusion that the use of monetary aggregates leads, at most, to marginal improvements in forecasting inflation and are of limited value in that context. With continual innovation in financial markets, the impact on the measurement of monetary aggregates will continue to present problems in the form of monetary control. Aggregates that have been adjusted to take account of financial innovations and the riskiness of the assets, along the lines proposed recently by Barnett and Wu [84], have the potential to make a valuable contribution to the future development of monetary policy and is likely to be a fruitful area for future research. Having established the preferred monetary aggregate construct for forecasting purposes, a next vital question to ask is how best to incorporate our chosen aggregate into a composite leading indicator of inflation for detecting early warning signs of upturns and downturns in

---

[8] We thank an anonymous reviewer for this suggestion.

[9] We also tried to break down the dominance of the large-scale common trend in monetary aggregates (in comparison to their smaller-scale profiles) by differencing the monetary aggregates and repeating the PCR approach on such detrended data. The models predicted the change in inflation rate six months ahead, and the predicted difference was then added to the current inflation rate. Even though the clear dominance of the averaging eigenvector was weakened, the prediction performance did not improve.

inflation, along the lines outlined in [85–87]. The construction of a composite leading indicator of US inflation along with a detailed timing classification of the cycles is recommended for future study.

How should we interpret our finding that monetary aggregates are not helpful in forecasting US inflation? There are several possible explanations for these findings. As [11] points out, since the mid 1980s, inflation has become more difficult to forecast, in the sense that it has become more difficult for an inflation forecaster to provide value added relative to a univariate model. This is not unique to monetary aggregates, but is a more general phenomenon [88]. Another explanation, more specific to money, is that monetary aggregates may be less informative when inflation is low and stable than during periods of high and volatile inflation. The reason is that one can think of money demand shocks as noise that obscures the signal from the monetary aggregates. Thus, in periods when inflation and money growth are both subdued, the signal to noise ratio of the monetary aggregates is likely to be low [89].

The "market is efficient" in the sense that it is difficult to beat the naïve RW model. The IORW measure is best suited for our purposes, because rather than worrying about the precise RMSE values, we should be concerned about by how much we can beat the rather obvious RW strategy. In other experiments (not reported here) we found that the longer the prediction horizon, the harder it is to accurately predict the inflation rate. It seems that inclusion of historical inflation rates alone works quite well for the KRLS model class. We may speculate that the value of money is implicitly represented in the inflation rates, and thus their explicit inclusion as input variables does not help; rather, it can actually make things worse because of the undersampling problems.

The validation set is just a sample; therefore, we cannot rely solely on picking the best model on the validation set. Slightly worse models on the validation set may be slightly better on the test set. Therefore we constructed flat ensembles of several best performing models on the validation set. We tested such ensembles on the test set, but their performance was inferior to that of the single winner picked on the validation set.

We also noted an interesting fact that the IORW measures did not change drastically with increasing prediction horizon. This indicates that, for longer prediction horizons, it is more difficult to predict the inflation rates, but the naïve model is less suited as well, so things cancel out.

The class of recurrent neural networks used for this problem domain appeared to have offered some advantages over simple recurrent networks. Although the additional feedback connections used increase the complexity of the model, the improvement in performance appears to outweigh the negative effect of increased number of parameters. Also, the success of the MRNs is largely dependent on the hyper-parameters used, most of which have to be empirically established. There is no universal rule to determine the level and nature of recurrent and self-recurrent connections. A key problem with the approach taken is the use of gradient-descent-based learning algorithms. The problems with back-propagation through time, for example, are well documented [77,79]. Bengio et al. [77] in particular has shown that the error gradient vanishes exponentially for dependent observations that are separated by intervening nonrelated observations, where a greater number of intervening observations yields poorer generalization performance. Although the MRN offers modifications to the architecture which may alleviate some of the problems posed by gradient-descent learning methods (i.e., through the formation of sluggish state spaces), we will also investigate other learning algorithms that are better able to find solutions the architecture is theoretically capable of representing (see [90–92]).

In the case of the kernel-based regression method, further investigation is required to understand and improve the stability of the model with respect to the addition of further input variables (e.g., measures of money). It may be that other compound measures of money may be more useful; however, this is likely to be model dependent. Further investigation into controlling the model complexity (e.g., determining the number of kernels) may provide additional improvement and insight into the efficacy of this approach to inflation forecasting. This will require more sophisticated versions of cross-validation than the one employed in this paper.

Future work will need to address the problem of why the kernel models break down when measures of money are added. It seems that this issue is related to over-fitting. When we tried using money alone, we could (after some work) get an excellent fit on the training set, but the models were inferior on the validation (and test) set. In other words, the learnt relationships by KRLS between past measures of money and future inflation were not representing anything real. Indeed, we experimented with another class of nonlinear predictors, namely evolutionary neural networks [93,94] (results not reported here). This framework can produce powerful predictors, but controlling the complexity in the evolutionary setting becomes more challenging. Most models tended to over-fit the training sample, leading to inaccurate out-of-sample forecasts.

We noted that, using KRLS, a good fit on the training set was obtained more naturally and easily when only past inflation rates were considered. This would suggest that, conditional on the kernel machine class used, there is little mutual information between past bulk measures of money and future inflation. Of course, future work will need to validate or refute this hypothesis.

Acknowledgements

Appendix. US monetary aggregates

Definition	Source/construction
M1	St. Louis Federal Reserve Bank, Fred database, USA
M2	St. Louis Federal Reserve Bank, Fred database, USA
MZM	St. Louis Federal Reserve Bank, Fred database, USA
M3	St. Louis Federal Reserve Bank, Fred database, USA
Divisia (MSI) M1	St. Louis Federal Reserve Bank, Fred database, USA
Divisia (MSI) M2	St. Louis Federal Reserve Bank, Fred database, USA
Divisia (MSI) MZM	St. Louis Federal Reserve Bank, Fred database, USA
Divisia (MSI) M3	St. Louis Federal Reserve Bank, Fred database, USA
Divisia (MSI) M1	Upper Envelope Benchmark Rate
Divisia (MSI) M2	Upper Envelope Benchmark Rate
Divisia (MSI) MZM	Upper Envelope Benchmark Rate
Divisia (MSI) M3	Upper Envelope Benchmark Rate
Divisia (MSI) M1	Baa Benchmark Rate
Divisia (MSI) M2	Baa Benchmark Rate
Divisia (MSI) MZM	Baa Benchmark Rate
Divisia (MSI) M3	Baa Benchmark Rate

References

[1] L. Bachmeier, S. Leelahanon, Q. Li, Money growth and inflation in the United States, Macroeconomic Dynamics 11 (2007) 113–127.
[2] S. Galluccio, J.-P. Bouchaud, M. Potters, Rational decisions, random matrices and spin glasses, Physica A: Statistical and Theoretical Physics 259 (3–4) (1998) 449–456.
[3] L. Laloux, P. Cizeau, J.-P. Bouchaud, M. Potters, Noise dressing of financial correlation matrices, Physical Review Letters 83 (7) (1999) 1467–1470.
[4] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A. Nunes Amaral, H. Eugene Stanley, Universal and nonuniversal properties of cross correlations in financial time series, Physical Review Letters 83 (7) (1999) 1471–1474.
[5] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A. Nunes Amaral, T. Guhr, H.E. Stanley, Random matrix approach to cross correlations in financial data, Physical Review E 65 (6) (2002).
[6] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A. Nunes Amaral, H. Eugene Stanley, Econophysics: financial time series from a statistical physics point of view, Physica A: Statistical Mechanics and its Applications 279 (1–4) (2000) 443–456.
[7] S. Saito, Verification of possibility of forecasting economical time series using neural network and digital filter, in: Third International Conference on Natural Computation, 3, 2007, pp. 381–388.
[8] S. Chen, L. Jani, C. Tai, Computational Economics: A Perspective from Computational Intelligence, Idea Publishing Group, Hershey PA, USA, 2006.
[9] European Central Bank, The outcome of the ECB's evaluation of its monetary policy strategy, ECB Monthly Bulletin (2003) 87–102. June.
[10] J. Stock, M. Watson, Forecasting inflation, Journal of Monetary Economics 44 (1999) 293–335.
[11] J. Stock, M. Watson, Why has US inflation become harder to forecast? Journal of Money, Credit, and Banking 39 (2007) 3–33.
[12] B. Bernanke, Monetary aggregates and monetary policy at the Federal Reserve: a historical perspective, in: Remarks at the fourth ECB central banking conference, Frankfurt, Germany, November 10, 2006, Board of Governors of the Federal Reserve System, Washington, DC, 2006.
[13] E. Nelson, Direct effects of base money on aggregate demand: theory and evidence, Journal of Monetary Economics 49 (4) (2002) 687–708.
[14] E. Nelson, The future of monetary aggregates in monetary policy analysis, Journal of Monetary Economics 50 (5) (2003) 1029–1059.
[15] E. Leeper, J. Rouch, Putting 'M' back in monetary policy, Journal of Money, Credit and Banking 35 (2003) 1217–1256.
[16] C. Carlstrom, T. Fuerst, Thinking about monetary policy without money, International Finance 7 (2) (2004) 325–347.
[17] M. King, No money, no inflation–the role of money in the economy, Bank of England Quarterly Bulletin (2002) 162–177. Summer.
[18] D. Schunk, The relative forecasting performance of the Divisia and simple sum monetary aggregates, Journal of Money, Credit, and Banking 33 (2001) 272–283.
[19] L. Drake, T. Mills, A new empirically weighted monetary aggregate for the United States, Economic Inquiry 43 (2005) 138–157.
[20] T. Elger, B. Jones, B. Nilsson, Forecasting with monetary aggregates: recent evidence for the United States, Journal of Economics and Business 15 (2006) 428–446.
[21] T. Simpson, The redefined monetary aggregates, Federal Reserve Bulletin (1980) 97–114. February.
[22] G. Moore, R. Porter, D. Small, Modeling the disaggregated demands for M2 and M1: the US experience in the 1980s, in: P. Hooper, K. Johnson, D. Kohn, D. Lindsey, R. Porter, R. Tryon (Eds.), Financial Sectors in Open Economies: Empirical Analysis and Policy Issues, Board of Governors of the Federal Reserve System, Washington, DC, 1990, pp. 21–105.
[23] J. Carlson, D. Hoffman, B. Keene, R. Rasche, Results of a study of the stability of cointegrating relations comprised of broad monetary aggregates, Journal of Monetary Economics 46 (2) (2000) 345–383.
[24] J. Duca, D. VanHoose, Recent developments in understanding the demand for money, Journal of Economics and Business 56 (2004) 247–272.
[25] R. Anderson, R. Rasche, Retail sweep programs and bank reserves: 1994–1999, Federal Reserve Bank of St. Louis *Review* 83 (2001) 51–72.
[26] B. Jones, D. Dutkowsky, T. Elger, Sweep programs and optimal monetary aggregation, Journal of Banking and Finance 29 (2) (2005) 483–508.
[27] D. Dutkowsky, B. Cynamon, B. Jones, US narrow money for the twenty-first century, Economic Inquiry 44 (1) (2006) 142–152.
[28] W.A. Barnett, Economic monetary aggregates: an application of index number and aggregation theory, Journal of Econometrics 14 (1) (1980) 11–48.
[29] W.A. Barnett, The optimal level of monetary aggregation, Journal of Money, Credit and Banking 14 (4) (1982) 687–710.
[30] L. Törnqvist, The Bank of Finland's consumption price index, Bank of Finland Bulletin 10 (1936) 1–8.
[31] H. Theil, Economics and Information Theory, North-Holland, Amsterdam, 1967 (Chapter 5).
[32] E. Diewert, Intertemporal consumer theory and the demand for durables, Econometrica 42 (3) (1974) 497–516.
[33] E. Diewert, Exact and superlative index numbers, Journal of Econometrics 4 (2) (1976) 115–145.
[34] W.A. Barnett, D. Fisher, A. Serletis, Consumer theory and the demand for money, Journal of Economic Literature 30 (1992) 86–119.
[35] R. Anderson, B. Jones, T. Nesmith, Monetary aggregation theory and statistical index numbers, Federal Reserve Bank of St. Louis *Review* 79 (1997) 31–51.
[36] L. Drake, A.W. Mullineux, J. Agung, One Divisia money for Europe, Applied Economics 29 (1997) 775–786.

[37] C. Hulten, Divisia index numbers, Econometrica 41 (6) (1973) 1017–1025.
[38] M. Hancock, Divisia money, Bank of England Quarterly Bulletin 45 (1) (2005) 39–46.
[39] H. Berument, N. Dincer, Inflation and inflation uncertainty in the G-7 countries, Physica A: Statistical Mechanics and its Applications 348 (2005) 371–379.
[40] H. Berument, E. Yuksel, Effects of adopting inflation targeting regimes on inflation variability, Physica A: Statistical Mechanics and its Applications 375 (2007) 265–273.
[41] J. Binner, A. Gazely, G. Kendall, Evaluating the performance of a EuroDivisia index using artificial intelligence techniques, in: 4th International Conference on Computational Intelligence in Economics and Finance, 2005, pp. 871–874.
[42] F. Mishkin, Headline versus core inflation in the conduct of monetary policy, Speech presented at the Business Cycles, International Transmission and Macroeconomic Policies Conference, HEC Montreal, Montreal, October. Board of Governors of the Federal Reserve System, Washington, DC, 2007.
[43] F. Mishkin, K. Schmidt-Hebbel, One decade of inflation targeting in the world: what do we know and what do we need to know? in: N. Loayza, R. Soto (Eds.), Inflation Targeting: Design, Performance, Challenges, Central Bank of Chile, Santiago, 2001, pp. 117–219.
[44] J.S. Greenlees, R.B. McClelland, Addressing misconceptions about the consumer price index, Monthly Labor Review 131 (8) (2008) 3–19.
[45] Todd. Clark, A comparison of the CPI and PCE price index, Federal Reserve Bank of Kansas City *Economic Review* (Third Quarter) (1999) 15–29.
[46] C. McCully, B. Moyer, K. Stewart, A reconciliation between the consumer price index and the personal consumption expenditures price index, mimeo, Bureau of Labor Statistics, September 2007, 2007.
[47] T. Garner, G. Janini, W. Passero, L. Paszkiewicz, M. Vendermia, The CE and the PCE: a comparison, Monthly Labor Review 129 (9) (2006) 20–46.
[48] E. Nakamura, Forecasting inflation using a neural network, Economics Letters 86 (2005) 373–378.
[49] E. Nakamura, Pass-through in retail and wholesale, American Economic Review 98 (2) (2008) 430–437.
[50] R. Anderson, B. Jones, T. Nesmith, Building new monetary services indexes: concepts, data, and methods, Federal Reserve Bank of St. Louis *Review* 79 (1997) 53–82.
[51] Arturo. Estrella, Why does the yield curve predict output and inflation? Economic Journal 115 (505) (2005) 722–744.
[52] A. Ang, M. Piazzesi, M. Wei, What does the yield curve tell us about GDP growth? Journal of Econometrics 131 (1–2) (2006) 359–403.
[53] J. Moody, Economic forecasting: challenges and neural network solutions, in: Paper presented at the International Symposium on Artificial Neural Networks, Hsinchu, Taiwan, December 1995, 1995.
[54] G. Zhang, B. Patuwo, M. Hu, Forecasting with artificial neural networks: the state of the art, International Journal of Forecasting 14 (1998) 35–62.
[55] C. Giles, S. Lawrence, A. Tsoi, Noisy time series prediction using recurrent neural networks and grammatical inference, Machine Learning 44 (2001) 161–183.
[56] B. Schölkopf, A. Smola, Learning with Kernels, MIT Press, 2002.
[57] Y. Engel, S. Mannor, R. Meir, The kernel recursive least-squares algorithm, IEEE Transactions on Signal Processing 52 (8) (2004) 2275–2285.
[58] D. Andrews, Nonparametric kernel estimation for semiparametric models, Econometric Theory 11 (1995) 560–596.
[59] R. Dorsey, Neural networks with Divisia money: better forecasts of future inflation, in: M. Belongia, J.M. Binner (Eds.), Divisia Monetary Aggregation: Theory and Practice, Palgrave, 2000.
[60] K. Hornik, M. Stinchcombe, H. White, Multi-layer feedforward networks are universal approximators, Neural Networks 2 (1989) 359–366.
[61] K. Judd, Numerical Methods in Economics, MIT Press, 1998.
[62] B.A. Pearlmutter, Gradient calculations for dynamic recurrent neural networks: a survey, IEEE Transactions on Neural Networks 6 (5) (1995) 1212–1228.
[63] D. Rumelhart, G. Hinton, R. Williams, Learning internal representations by error propagation, in: Parallel Distributed Processing, Explorations in the Microstructure of Cognition, Vol. 1, Princeton University Press, 1986, pp. 318–362.
[64] S. Moshiri, N.E. Cameron, D. Scuse, Static, dynamic and hybrid neural networks in forecasting inflation, Computational Economics 14 (1999) 219–235.
[65] P. Tenti, Forecasting foreign exchange rates using recurrent neural networks, Applied Artificial Intelligence 10 (1996) 567–581.
[66] C. Ulbricht, Multi-Recurrent networks for traffic forecasting, in: Proceedings of the Twelfth National Conference in Artificial Intelligence, AAAI Press/MIT Press, Cambridge, MA, 1994, pp. 883–888.
[67] J. Binner, T. Elger, B. Nilsson, J. Tepper, Tools for non-linear time series forecasting in economics: an empirical comparison of regime switching vector autoregressive models and recurrent neural networks, Advances in Econometrics: Applications of Artificial Intelligence in Finance and Economics 19 (2004) 71–91.
[68] M. Jordan, Attractor dynamics and parallelism in a connectionist sequential machine, in: J. Diederich (Ed.), Artificial Neural Networks: Concept-Learning, IEEE Press, Piscataway NJ, 1990, pp. 112–127. Also, Proceedings of the 8th Annual Conference of the Cognitive Science Society, 1986, pp. 531–545.
[69] J. Stock, M. Watson, A comparison of linear and non-linear univariate models for forecasting macroeconomic time series, in: R. Engle, H. White (Eds.), Cointegration, Causality and Forecasting: A Festschrift in Honor of Clive Granger, Oxford University Press, 1999.
[70] J. Binner, T. Elger, B. Nilsson, J. Tepper, Predictable non-linearities in US inflation, Economic Letters 93 (3) (2006) 323–328.
[71] G. Dorffner, Neural networks for time series processing, Neural Network World 6 (4) (1996) 447–468.
[72] J. Elman, Finding structure in time, Cognitive Science 14 (1990) 179–211.
[73] P.J. Werbos, Backpropagation through time: what it does and how to do it, in: Proceedings of IEEE Special Issue on Neural Networks, vol. 78, 1990, pp. 1550–1560.
[74] R.J. Williams, J. Peng, An efficient gradient-based algorithm for on-line training of recurrent network trajectories, Neural Computation 2 (1990) 490–501.
[75] C. Darken, J. Chang, J. Moody, Learning rate schedules for faster stochastic gradient search, in: Proceedings of the IEEE Workshop on Neural Networks for Signal Processing, vol. 2, IEEE Press, 1992.
[76] S. Lawrence, C. Giles, S. Fong, Natural language grammatical inference with recurrent neural networks, IEEE Transactions on Knowledge and Data Engineering 12 (1) (2000) 126–140.
[77] Y. Bengio, P. Simard, P. Frasconi, Learning long-term dependencies with gradient descent is difficult, IEEE Transactions on Neural Networks 5 (2) (1994) 157–166.
[78] N. Sharkey, A. Sharkey, S. Jackson, Are SRN's sufficient for modeling language acquisition? in: P. Broeder, J. Murre (Eds.), Models of Language Acquisition: Inductive and Deductive Approaches, Oxford University Press, 2000, pp. 33–54.
[79] S. Haykin, Neural Networks: A Comprehensive Foundation, 2nd edition, Prentice Hall, Upper Saddle River, NJ, 1999.
[80] K. Kim, Financial time series forecasting using support vector machines, Neurocomputing 55 (1–2) (2003) 307–319.
[81] L. Wang, J. Zhu, Financial market forecasting using a two-step kernel learning method for the support vector regression, Annals of Operations Research 174 (2010) 103–120.
[82] R. Rosipal, L.J. Trejo, Kernel partial least squares regression in reproducing kernel Hilbert space, Journal of Machine Learning Research 2 (2002) 97–123.
[83] I.T. Jolliffe, A note on the use of principal components in regression, Journal of the Royal Statistical Society 31 (2) (1982) 300–303. Series C.
[84] W.A. Barnett, S. Wu, On user costs of risky monetary assets, Annals of Finance 1 (1) (2005) 35–50.
[85] J. Binner, A. Fielding, A. Mullineux, Divisia money in a composite leading indicator of inflation, Applied Economics 31 (1999) 1021–1031.
[86] J. Binner, R. Bissoondeeal, A. Mullineux, A composite leading indicator of the inflation cycle for the Euro area, Applied Economics 37 (2005) 1257–1266.
[87] J. Binner, S. Wattam, A new composite leading indicator of inflation for the UK: a Kalman filter approach, Global Business and Economics Review 5 (2) (2003) 242–264.
[88] A. Atkeson, L. Ohanian, Are Phillips curves useful for forecasting inflation? Federal Reserve Bank of Minneapolis Quarterly Review 25 (1) (2001) 2–11.
[89] A. Estrella, F. Mishkin, Is there a role for monetary aggregates in the conduct of monetary policy? Journal of Monetary Economics 40 (1997) 279–304.
[90] H.T. Siegelmann, E.D. Sontag, Turing computability with neural nets, Applied Mathematics Letters 4 (1991) 77–80.
[91] F. Gers, D. Eck, J. Schmidhuber, Applying LSTM to time series predictable through time-window approaches, in: International Conference on Artificial Neural Networks, Vienna, Austria, 2001, pp. 669–675.
[92] H. Jaeger, H. Haas, Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication, Science 2 (304(5667)) (2004) 78–80.
[93] X. Yao, Evolving artificial neural networks, Proceedings of the IEEE 87 (9) (1999) 1423–1447.
[94] K. Stanley, R. Miikkulainen, Evolution of neural networks through augmenting topologies, Evolutionary Computation 10 (2002) 99–127.

Further reading

[1] C. Andreeski, P. Vasant, Comparative analysis on time series with included structural break, in: Power Control and Optimization: Proceedings of the Second Global Conference on Power Control and Optimization, American Institute of Physics Conference Proceedings series 1159 (1) (2009) 2009.

[2] D. Andrews, Asymptotics for Semi-parametric econometric models via stochastic equicontinuity, Econometrica 62 (1994) 43–72.

[3] Hans J. Bierens, Uniform consistency of kernel estimators of a regression function under generalized conditions, Journal of the American Statistical Association 78 (383) (1983) 699–707.

[4] J. Binner, G. Kendall, A. Gazely, Evolving neural networks with evolutionary strategies: a new application to Divisia money, Advances in Econometrics: Applications of Artificial Intelligence in Finance and Economics 19 (2004) 127–143.

[5] J. Duca, Financial technology shocks and the case of the missing M2, Journal of Money, Credit, and Banking 32 (4) (2000) 820–839.

[6] William H. Greene, Econometric Analysis, 6th edition, Pearson-Prentice Hall, 2008.

[7] Q. Li, J.S. Racine, Nonparametric Econometrics: Theory and Practice, Princeton University Press, 2007.

[8] Z. Michalewicz, D. Fogel, How to Solve It, Springer-Verlag, ISBN: 3-540-66061-5, 2000.