CAUSAL MODELING WITH APPLICATIONS TO THE FOREIGN EXCHANGE MARKET

A Dissertation

by

BRIAN DAVID DEATON

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	David A. Bessler
Committee Members,	David J. Leatham
	James W. Mjelde
	Willa W. Chen
Head of Department,	C. Parr Rosson, III

December 2013

Major Subject: Agricultural Economics

ABSTRACT

A combination of time series models and causal search algorithms is applied to the foreign exchange markets to find causal linkages between the six most widely traded currencies (Australian dollar, Canadian dollar, euro, Great Britain pound sterling, Japanese yen, and United States dollar). This information is used in portfolio management to improve risk management, to visualize the causal connections between currencies, and enhance the forecasting ability of time series models.

In the first section, a method is presented that decomposes portfolio risk so that risk contributions sum to the total portfolio's risk. This decomposition is based upon a market's underlying independent risk factors, which are found empirically using a causal search algorithm based on independent component analysis. In an application, independent risk contributions are constrained during portfolio optimizations, and the internal risk characteristics of the resulting portfolios are shown to be superior to those constructed using more traditional constraints.

In the second section, three causal search algorithms are used to identify causal relationships amongst the six most widely traded currencies in the years 2009-2011. The intent is to discover causal relationships within each year and to observe how these causal relationships change over time. The causal relationships are presented as directed acyclic graphs, and these are relatively stable over time. There might be, however, latent variables that affect the six most widely traded currencies.

In the third section, probability forecasts of the Swiss franc/euro (CHF/EUR) exchange rate from three different time series models are generated before, surrounding, and after the placement of a floor on the CHF/EUR exchange rate by the Swiss National Bank. The goal is to determine whether the exchange rate floor has a positive, negative, or insignificant affect on the calibration of the probability forecasts. Forecasts from the models are ranked with score metrics, and a graphical d-separation criterion is used in an attempt to identify the preferred model based on forecast performance. The study finds evidence that the floor on the CHF/EUR has a negative impact on the forecasting performance of the three time series models.

ACKNOWLEDGEMENTS

Thank you to my father for his encouragement and unwavering support, to my committee chair, Dr. Bessler, for his inspiration and patience, and to my committee members, Dr. Leatham, Dr. Mjelde, and Dr. Chen, for their guidance and support throughout the course of this research. Thanks also to my friends, colleagues, and the agricultural economics department faculty and staff for making my time at Texas A&M University a great experience.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

TABLE Page

CHAPTER I

INTRODUCTION

Three applications of the combination of time series models and causal search algorithms to the foreign exchange market are contained in Chapters II-IV of this dissertation. The objective of Chapter II is to show that portfolio risk can be decomposed so that independent risk contributions sum to the total portfolio's risk. The internal risk characteristics of portfolios can be improved by placing constraints on these independent risk contributions during portfolio optimization. In Chapter III, the causal structure of the foreign exchange market is found in each of the years 2009, 2010, and 2011. The intent of this chapter is to discover the causal relationships between the six most widely traded currencies and to observe how these causal relationships change over time. The objective of Chapter IV is to determine whether a floor placed on the Swiss franc/euro (CHF/EUR) exchange rate has a positive, negative, or insignificant affect on the calibration of probability forecasts from three time series models. Probability forecast calibration is measured in periods of time before, surrounding, and after the placement of the floor on the CHF/EUR exchange rate. Scoring metrics and a method based on directed acyclic graphs are used in an attempt to identify the preferred forecasting system.

A method for decomposing portfolio risk based on a portfolio's underlying independent risk factors is presented in Chapter II. This form of portfolio risk decomposition allows a portfolio manager to create a risk budget for independent risk

factors because the independent risk contributions sum to the total portfolio's risk. In contrast, risk budgeting is impossible with traditional marginal portfolio risk decomposition because marginal risk contributions are relative to an existing portfolio and can be negative.

In a sample application of this technique, a portfolio's underlying independent risk factors are found empirically using methods based on independent component analysis. The mean-variance model is used for portfolio optimization and its risk metric, the portfolio variance, is decomposed into its independent risk contributions. The independent risk contributions are constrained during the optimization process to increase portfolio diversification.

The portfolio constructed with constraints based on risk contributions of the independent factors is compared to five other portfolios constructed using the canonical mean-variance model with constraints on the following: relative portfolio weights, independent factor relative portfolio weights, percentage marginal risk contributions, and independent factor percentage marginal risk contributions. An efficient frontier is generated for each of the portfolios and a comparison is made of the loss of efficiency due to each of the constraint sets. Additionally, the performance and internal characteristics of each of these portfolios are analyzed on a small out-of-sample data set.

Three different causal search algorithms (PC, FCI, and LiNGAM) in conjunction with vector autoregression (VAR) time series models are used in Chapter III to find possible causal structures underlying the foreign currency market in the years 2009, 2010, and 2011. This modeling process results in a directed acyclic graph that illustrates

the contemporaneous causal linkages between currencies. The three causal search algorithms do not necessarily find the same causal structure when using the same data set.

The causal structure underlying the six most widely traded currencies (Australian dollar, Canadian dollar, euro, Great Britain pound sterling, Japanese yen, and United States dollar) on a 15 minute time frame is found for each year (2009, 2010, and 2011) by each algorithm. The stability over time of the sequence of causal structures from each algorithm is assessed, and the three causal structures found by the three algorithms within each year are compared to each other.

Chapter IV assesses the impact of the Swiss National Bank's floor on the Swiss franc/euro (CHF/EUR) exchange rate in 2011-2012 on the probability forecasts of the CHF/EUR from three time series models. The goal of the study is to determine the effect of the CHF/EUR floor on the calibration of the probability forecasts. A ranking of the models based on the accuracy of their forecasts is also considered.

One-step-ahead forecasts of the CHF/EUR probability distribution are based on a series of intraday futures data for the Australian dollar, Canadian dollar, euro, Great Britain pound sterling, Japanese yen, and the United States dollar all in terms of the Swiss franc. Probability forecasts of the CHF/EUR are generated from a vector autoregression model, a VAR model augmented with the LiNGAM causal learning algorithm, and a univariate autoregressive model built on the independent components of an independent component analysis.

The forecasted probability distributions are tested for calibration and ranked with the mean-squared error and the probability score metrics in periods of time before, surrounding, after, and long after the beginning of the floor on the CHF/EUR exchange rate. A method based on directed acyclic graphs is used as a complement to the scoring metrics in an attempt to identify whether or not the preferred forecasting system changes over time.

CHAPTER II

RISK ATTRIBUTION WITH INDEPENDENT FACTORS

Introduction

The risk attribution literature is based on a marginal form of portfolio risk decomposition. One disadvantage of a marginal analysis is that a portfolio manager cannot use it to create a risk budget that sums to the total portfolio risk. Marginal risk contributions are relative to an existing portfolio and can be negative, which makes risk budgeting impossible.

The primary objective of this chapter is to show how portfolio risk can be decomposed so that risk contributions sum to the total portfolio's risk. In an application, it is shown that the internal risk characteristics of portfolios can be improved by placing constraints on these risk contributions. The risk decomposition is based upon a market's underlying independent risk factors, which are found empirically using methods based on independent component analysis.

The analysis is built on the framework of Meucci (2007) for analyzing risk contributions from a set of arbitrary risk factors. In this framework, a portfolio's profit and loss Π is represented by the product of a vector of risk factors $\boldsymbol{F}$ and a vector of their respective exposures $\boldsymbol{b}$

$$\Pi = \boldsymbol{b}^T \boldsymbol{F} \tag{2.1}$$

where the superscript *T* denotes the transpose of a vector or matrix. The vectors $\boldsymbol{b}$ and $\boldsymbol{F}$ can be defined in a variety of ways including the case where $\boldsymbol{b} \equiv \boldsymbol{\omega}$ is a vector of relative portfolio weights and $\boldsymbol{F} \equiv \boldsymbol{R}$ is a vector of security returns.

It is often desirable to analyze a portfolio in terms of a set of new risk factors $\widetilde{\boldsymbol{F}}$ that completely span the market. In this case, there is an invertible matrix $\boldsymbol{P}$ that linearly transforms the set of original risk factors into the set of new risk factors

$$\widetilde{\boldsymbol{F}} \equiv \boldsymbol{P}\boldsymbol{F}. \tag{2.2}$$

The matrix $\boldsymbol{P}$ can be interpreted as a "pick" matrix containing rows that transform each original factor into its respective new factor. The exposures to the new factors $\widetilde{\boldsymbol{b}}$ can then be computed in terms of the exposures to the original factors $\boldsymbol{b}$ as follows

$$\Pi = \mathbf{b}^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{P}\mathbf{F} = \tilde{\mathbf{b}}^{\mathrm{T}}\tilde{\mathbf{F}} \tag{2.3}$$

so that

$$\widetilde{\boldsymbol{b}} \equiv \boldsymbol{P}^{-T}\boldsymbol{b} \tag{2.4}$$

where the superscript -1 denotes the inverse of a matrix.

Portfolio Theory

For simplicity, only mean-variance analysis will be considered here. The methods presented below can easily be used with expected utility optimization or with alternative risk and reward measures. In the mean-variance portfolio theory of Markowitz (1952), an investor only cares about the first two moments of the distribution of returns. The goal of the investor is to create a portfolio of risky assets that maximizes

his expected return for a given level of risk, which is measured by the variance or the standard deviation of the portfolio return.

The risk factors $\boldsymbol{F}$ of the previous section are defined as the returns $\boldsymbol{R}$ on a set of N securities

$$\boldsymbol{R} = \begin{pmatrix} R_1 \\ \vdots \\ R_N \end{pmatrix}. \tag{2.5}$$

where each R_i is the return on security i. Canonical mean-variance optimization is performed for an investment horizon of one period so that a reference to time is unnecessary. The expected returns for the investment horizon are

$$E(\boldsymbol{R}) = \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_N \end{pmatrix} \tag{2.6}$$

where each μ_i is the expected return of security i, and the expected returns have covariance matrix

$$Cov(\boldsymbol{R}) = \boldsymbol{\Omega}. \tag{2.7}$$

The exposures $\boldsymbol{b}$ to the set of returns $\boldsymbol{R}$ are defined as a set of relative portfolio weights $\boldsymbol{\omega}$. A portfolio is a weighted average of N risk factors with relative weight ω_i assigned to factor i

$$\boldsymbol{\omega} = \begin{pmatrix} \omega_1 \\ \vdots \\ \omega_N \end{pmatrix}. \tag{2.8}$$

Relative portfolio weights can be either positive for a long position or negative for a short position but must sum to one: $\sum_{i=1}^{N} \omega_i = 1$. Most portfolios also include a risk free

security with return R_{rf} and relative portfolio weight ω_{rf}. In this case, the portfolio's expected return is

$$R_P = \omega_{rf} R_{rf} + \boldsymbol{\omega}^T \boldsymbol{\mu} = \omega_{rf} R_{rf} + \sum_{i=1}^{N} \omega_i \mu_i. \tag{2.9}$$

The variance of the portfolio's return is

$$\sigma_P^2 = \boldsymbol{\omega}^T \boldsymbol{\Omega} \boldsymbol{\omega} = \sum_{i=1}^{N} \sum_{j=1}^{N} \omega_i \, \omega_j \Omega_{i,j} \tag{2.10}$$

where $\Omega_{i,j}$ is the covariance between factor i and factor j.

The canonical mean-variance model is

$$Minimize \; \lambda \sigma_P^2 - (1 - \lambda) R_P \tag{2.11}$$

$$subject \; to \;\; \omega_{rf} + \sum_{i=1}^{N} \omega_i = 1 \tag{2.12}$$

where λ is a risk aversion parameter; a risk neutral investor has $\lambda = 0$ while a highly risk averse investor has $\lambda = 1$ (Zenios 2007). This model can be used to trace an efficient frontier by varying λ across the interval [0, 1]; every λ in this interval corresponds to an efficient portfolio that has minimum variance for a given expected return.

Additional constraints can be added to the mean-variance model to achieve certain objectives. It is well known that the mean-variance model tends to allocate too much relative weight to certain assets (Green 1992; Gerard and Tutuncu 2007). Thus, a common addition is to put constraints on the relative portfolio weights so that the portfolio wealth is not focused on any one asset

$$\omega_i \leq m \quad for\ i = 1, \dots, N. \tag{2.13}$$

This idea can be generalized so that limits are placed on arbitrary factor exposures $\boldsymbol{b}$

$$b_i \leq m \quad for\ i = 1, \dots, N. \tag{2.14}$$

Marginal Risk Contributions

In the risk attribution literature, statistical measures of risk are decomposed by using the fact that they are homogeneous of degree one. For instance, popular risk measures such as standard deviation, value at risk (VaR), and expected shortfall (ES) all share this property. When the portfolio exposures $\boldsymbol{b}$ are scaled by a factor *k,* a risk measure $\mathcal{R}$ that is homogeneous of degree one increases by that same factor

$$\mathcal{R}(k\boldsymbol{b}) = k * \mathcal{R}(\boldsymbol{b}). \tag{2.15}$$

Risk measures that behave this way can be decomposed into a summation by taking the derivative of each side of the above equation with respect to k

$$\sum_{i=1}^{N} b_i \frac{\partial \mathcal{R}(k\boldsymbol{b})}{\partial b_i} = \mathcal{R}(\boldsymbol{b}). \tag{2.16}$$

Now set $k = 1$ to obtain

$$\mathcal{R}(\boldsymbol{b}) = \sum_{i=1}^{N} b_i \frac{\partial \mathcal{R}(\boldsymbol{b})}{\partial b_i}. \tag{2.17}$$

This equation states that the risk of the portfolio, which is a scalar quantity, is a sum of the risk contributions from each individual factor. The risk from each individual factor is expressed as a product of the factor exposure b_i and the marginal rate of change in risk

per unit change in the factor exposure $\frac{\partial \mathcal{R}(\boldsymbol{b})}{\partial b_i}$. Each risk contribution can also be written in terms of a percentage contribution to portfolio risk as follows

$$\frac{b_i * \frac{\partial \mathcal{R}(\boldsymbol{b})}{\partial b_i}}{\mathcal{R}(\boldsymbol{b})} * 100 \tag{2.18}$$

so that the sum of the percentage contributions is 100.

As Litterman (1996) notes, this is a marginal analysis and useful only in analyzing risk relative to an existing portfolio. For example, eliminating a position that accounts for half the risk will not reduce the portfolio's risk by half. As a position is reduced in size, its marginal contribution to risk will also be reduced. Additionally, some positions have a negative marginal contribution to risk so that increasing the size of such a position will reduce the risk of the portfolio at the margin.

It is straightforward to obtain a risk metric decomposition in terms of a new set of factors via the risk decomposition computations that use the original set of factors. The marginal rates of change in risk per unit change in the new factor exposures are linear translations of the originals (Meucci 2007)

$$\frac{\partial \mathcal{R}(\tilde{\boldsymbol{b}})}{\partial \tilde{\boldsymbol{b}}} = \boldsymbol{P} \frac{\partial \mathcal{R}(\boldsymbol{b})}{\partial \boldsymbol{b}} \tag{2.19}$$

so that the risk measure in terms of the new factor exposures is

$$\mathcal{R}(\tilde{\boldsymbol{b}}) = \sum_{i=1}^{N} \tilde{b}_i \frac{\partial \mathcal{R}(\tilde{\boldsymbol{b}})}{\partial \tilde{b}_i}. \tag{2.20}$$

For mean-variance analysis, the portfolio's standard deviation will be used as the measure of risk

$$\mathcal{R}(\boldsymbol{b}) = \sqrt{\boldsymbol{b}^T Cov\{\boldsymbol{F}\}\boldsymbol{b}}, \tag{2.21}$$

where $Cov\{\boldsymbol{F}\}$ is the risk factor covariance matrix. The partial derivatives of the portfolio standard deviation are the following

$$\frac{\partial \mathcal{R}(\boldsymbol{b})}{\partial \boldsymbol{b}} = \frac{Cov\{\boldsymbol{F}\}\boldsymbol{b}}{\sqrt{\boldsymbol{b}^T Cov\{\boldsymbol{F}\}\boldsymbol{b}}}. \tag{2.22}$$

In the application below, $\boldsymbol{b}$ is the vector of relative weights on the original currencies $\boldsymbol{b} \equiv \boldsymbol{\omega}$ (with $\boldsymbol{F} \equiv \boldsymbol{R}$), and $\widetilde{\boldsymbol{b}}$ is the vector of relative weights on the independent factors $\widetilde{\boldsymbol{b}} \equiv \widetilde{\boldsymbol{\omega}}$ (with $\widetilde{\boldsymbol{F}} \equiv \widetilde{\boldsymbol{R}}$).

Total Risk Contributions

Many risk measures, such as the portfolio standard deviation, are functions of statistics that can be decomposed into a summation if the factors $\boldsymbol{b}$ are independent

$$\mathcal{R}(\boldsymbol{b}) = \sum_{i=1}^{N} \mathcal{R}(b_i). \tag{2.23}$$

This decomposition will be referred to as a "total risk decomposition" to distinguish it from the marginal risk decomposition described earlier. In percentage form each total risk contribution is

$$\frac{\mathcal{R}(b_i)}{\mathcal{R}(\boldsymbol{b})} * 100 \tag{2.24}$$

and the sum of the percentage contributions is 100.

The total risk decomposition is simpler to compute and easier to interpret than the marginal risk decomposition. Each total risk contribution is non-negative and can be interpreted as having a potentially adverse effect on the portfolio (as opposed to the

negative contributions of the marginal decomposition). Furthermore, the sum of the total risk contributions is the risk of the portfolio. This allows portfolio managers to create a risk budget for independent factors that sums to 100% of their portfolio's risk. In this analysis, risk contributions are no longer relative to an existing portfolio, and it is no longer necessary to resort to interpreting risk contributions as potential loss contributions (as in Qian (2006)).

For the mean-variance model, a total risk decomposition of the variance is useful. Because the variance of a sum of independent variables is the sum of the variances, the total risk decomposition of a portfolio's variance is the following

$$\mathcal{R}(\tilde{\boldsymbol{b}}) = \sum_{i=1}^{N} \tilde{b}_i^2 Var\{\tilde{F}_i\}. \tag{2.25}$$

Independent Component Analysis

One way to operationalize the total risk decomposition is to use independent component analysis (ICA). Independent component analysis is based on the premise that N observed variables $x_1, \dots, x_N$ are linear combinations of underlying, statistically mutually independent source variables $s_1, \dots, s_N$

$$x_i = a_{i1}s_1 + a_{i2}s_2 + \cdots + a_{iN}s_N \quad for\ all\ \ i = 1, \dots, N \tag{2.26}$$

where the a_{ij} are mixing coefficients. This basic ICA model is written in vector-matrix form as

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s}. \tag{2.27}$$

The observed variables $\boldsymbol{x}$ are used to estimate both the unknown mixing coefficient matrix $\boldsymbol{A}$ and the unobserved independent components $\boldsymbol{s}$. The observed variables $\boldsymbol{x}$ and the independent components $\boldsymbol{s}$ are both assumed to have zero mean. If this assumption does not hold then the original observed variables, denoted by $\boldsymbol{x}_o$, can be centered in a preprocessing step

$$\boldsymbol{x} = \boldsymbol{x}_o - E(\boldsymbol{x}_o) \tag{2.28}$$

This preprocessing also forces the independent components to have zero mean since

$$E(\boldsymbol{s}) = \boldsymbol{A}^{-1}E(\boldsymbol{x}). \tag{2.29}$$

Estimation of the basic ICA model relies on the following assumptions (Hyvarinen, Karhunen, and Oja 2001)

1) The independent components are assumed to be statistically independent, but in application this does not need to be exactly true.
2) The mixing matrix $\boldsymbol{A}$ is assumed to be square and invertible for the sake of convenience and simplicity.
3) The independent components must have non-Gaussian distributions.

Some ICA models are slightly different from the basic ICA model and have their own assumptions; for details see Hyvarinen, Karhunen, and Oja (2001).

The independent components $\boldsymbol{s}$ are more than just uncorrelated; they are as statistically independent as possible. Achieving this requires more information than what is available in a correlation matrix unless all of the variables are normally distributed, in which case zero correlation is equivalent to independence. The estimation of

independent components uses higher order moments or other information such as the autocovariance structure for time series variables in addition to correlation information.

It is always possible to linearly transform the observed random variables $\boldsymbol{x}$ into uncorrelated variables. It is also often desirable that the transformed variables have variances equal to unity. The process called whitening transforms zero mean variables $\boldsymbol{x}$ into uncorrelated variables $\boldsymbol{z}$ that have unit variances. The result of whitening is a matrix $\boldsymbol{V}$ that decorrelates the observed data vector

$$\boldsymbol{z} = \boldsymbol{V}\boldsymbol{x}. \tag{2.30}$$

The matrix $\boldsymbol{V}$ is computed as

$$\boldsymbol{V} = \boldsymbol{D}^{-1/2}\boldsymbol{E}^{T} \tag{2.31}$$

where $\boldsymbol{E} = (\boldsymbol{e}_1 \dots \boldsymbol{e}_N)$ is the matrix whose columns are the unit-norm eigenvectors of the covariance matrix $\boldsymbol{C}_{\boldsymbol{x}} = E\{\boldsymbol{x}\boldsymbol{x}^T\}$ and $\boldsymbol{D} = diag(d_1 \dots d_N)$ is the diagonal matrix of the eigenvalues of $\boldsymbol{C}_{\boldsymbol{x}}$. There are an infinite number of matrices $\boldsymbol{V}$ that can create decorrelated components. This is the reason that estimation of the basic ICA model requires the higher order moments of non-Gaussian distributions (Hyvarinen, Karhunen, and Oja 2001).

The basic ICA model in equation 2.27 has the following ambiguities

1) The variances of the independent components cannot be determined.
2) The order of the independent components cannot be determined.

The first ambiguity occurs because any scalar multiple of one of the independent components s_i can be cancelled by dividing the corresponding column of the mixing

matrix $\boldsymbol{A}$ by that same multiple. The second ambiguity follows from modifying the model with any permutation matrix $\boldsymbol{P}$ and its inverse

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{P}^{-1}\boldsymbol{P}\boldsymbol{s}. \tag{2.32}$$

Now $\boldsymbol{A}\boldsymbol{P}^{-1}$ is the unknown mixing matrix and $\boldsymbol{P}\boldsymbol{s}$ are the independent components in a different order. In applications that use ICA, neither of these ambiguities is important; for an explanation, see Hyvarinen, Karhunen, and Oja (2001).

ICA Time Series

If the independent components are time series, as opposed to independent random variables in the basic ICA model, then the ICA model takes the following form

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t), \quad t = 1, \dots, T \tag{2.33}$$

where t is the time index (Hyvarinen, Karhunen, and Oja 2001). Since time series variables have more structure than independent random variables, the time series autocovariances may be used for estimation instead of the higher-order information that is required in the basic ICA model.

The AMUSE algorithm provides one method to estimate the model above (Hyvarinen, Karhunen, and Oja 2001). This algorithm requires the time-lagged covariance matrix in place of the higher-order moments used in the basic ICA model. The time-lagged covariance matrix is computed as

$$\boldsymbol{C}_{\tau}^{\boldsymbol{x}} = E\{\boldsymbol{x}(t)\boldsymbol{x}(t-\tau)^T\} \tag{2.34}$$

where τ is a lag constant, $\tau = 1, 2, 3, \dots$. This matrix contains the autocovariances of each time series and the covariances between the time series.

The algorithm is based on the fact that the instantaneous and lagged covariances of $\boldsymbol{s}(t)$ are zero due to independence. Hence, the time-lagged covariance matrix is used to find a matrix $\boldsymbol{B}$ so that all of the instantaneous and lagged covariances of

$$\boldsymbol{y}(t) = \boldsymbol{B}\boldsymbol{x}(t) \tag{2.35}$$

are equal to zero.

The AMUSE algorithm assumes that all of the independent components have autocovariances different from zero and different from each other. This assumption replaces the assumption of the basic ICA model that the independent components must have non-Gaussian distributions.

The AMUSE algorithm uses whitened, zero mean data $\boldsymbol{z}(t)$ as input and generates the separating matrix $\boldsymbol{W}$ as output so that

$$\boldsymbol{W}\boldsymbol{z}(t) = \boldsymbol{s}(t) \tag{2.36}$$

$$\boldsymbol{W}\boldsymbol{z}(t-\tau) = \boldsymbol{s}(t-\tau). \tag{2.37}$$

The time-lagged covariance matrix is modified to be symmetric by the following computation

$$\overline{\boldsymbol{C}}_{\tau}^{\boldsymbol{z}} = \frac{1}{2}[\boldsymbol{C}_{\tau}^{\boldsymbol{z}} + (\boldsymbol{C}_{\tau}^{\boldsymbol{z}})^{T}] \tag{2.38}$$

so that an eigenvalue decomposition on this new symmetric matrix is well defined.

The steps of the AMUSE algorithm are as follows (Hyvarinen, Karhunen, and Oja 2001)

1) Center and whiten the observed data $\boldsymbol{x}(t)$ to obtain $\boldsymbol{z}(t)$.
2) Compute the eigenvalue decomposition of the symmetric, time-lagged covariance matrix (equation 2.38) for some time lag τ.

3) The rows of the estimated separating matrix $\widehat{\boldsymbol{W}}$ are given by the eigenvectors.
4) The estimated separating matrix for the unwhitened data $\boldsymbol{x}$ is $\widehat{\boldsymbol{B}} = \widehat{\boldsymbol{W}}\boldsymbol{V}$ where $\boldsymbol{V}$ is defined in equation 2.31.

The separating matrix $\boldsymbol{B}$ can now be used in the risk contribution framework of Meucci (2007). Simply define the pick matrix $\boldsymbol{P}$ of equations 2.2 through 2.4 as the estimated separating matrix $\widehat{\boldsymbol{B}}$ to find a portfolio's exposure to the independent components.

Forecasting with independent components can be performed in a way similar to that described in Popescu (2009). Specifically,

1) Compute the independent components using the estimated separating matrix:

$$\boldsymbol{s}(t) = \widehat{\boldsymbol{B}}\boldsymbol{x}(t), \quad t = 1, \dots, T \tag{2.39}$$

2) Produce a forecast for each independent component using a time series model such as an autoregressive (AR) model.
3) Transform the forecasts of the independent components into forecasts of the original variables:

$$\boldsymbol{x}(t) = \widehat{\boldsymbol{A}}\boldsymbol{s}(t), \quad t = 1, \dots, T. \tag{2.40}$$

Dynamic Directed Graph Discovery (VAR-LiNGAM)

Another way to find independent risk factors is to use a causal discovery algorithm that incorporates ICA internally. The method used here, which will be called VAR-LiNGAM, is an enhancement of a vector autoregressive model with the LiNGAM causal discovery algorithm (Hyvarinen 2008; Hyvarinen et al. 2010; Shimizu et al.

2006). The LiNGAM algorithm uses ICA to find the matrix that transforms a vector of returns into its vector of corresponding independent factors.

VAR-LiNGAM is well suited for use in economics because it works with time series and, importantly, provides parameter estimates. This method is purportedly the first to fully identify a structural vector autoregressive model with the assumptions that there is acyclicity and that there are no latent variables. It uses the non-Gaussian structure of the data whereas other methods use only the covariance information, which is not always sufficient for identification.

Before the VAR-LiNGAM model is introduced, the two components on which it is built are discussed. The first component, a structural equation model (SEM), typically assumes that the observed data is independent and identically distributed; an SEM model does not consider the time series structure in data. A vector of contemporaneous returns $\boldsymbol{R}$ is modeled in SEM form as

$$\boldsymbol{R} = \boldsymbol{B}\boldsymbol{R} + \boldsymbol{e} \tag{2.41}$$

where $\boldsymbol{e}$ is a vector of disturbances and $\boldsymbol{B}$ is a matrix of coefficients; the diagonal of $\boldsymbol{B}$ is defined to be zero. Equation 2.41 can be easily transformed into an ICA model

$$\boldsymbol{R} = (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{e}. \tag{2.42}$$

The VAR-LiNGAM assumptions on the disturbance terms $\boldsymbol{e}$, enumerated below, allow ICA to be used as part of its estimation procedure.

The second component of VAR-LiNGAM, an autoregressive model, does consider the time series structure in data. The notation of a vector of return observations as a time series is

$$\boldsymbol{R}(t) = \begin{pmatrix} R_1(t) \\ \vdots \\ R_N(t) \end{pmatrix}, \qquad t \in 1, \dots, T \tag{2.43}$$

where $R_i(t)$ is the return on a particular asset $i \in 1, \dots, N$ at time $t \in 1, \dots, T$. An autoregressive model of multivariate time series return data $\boldsymbol{R}(t)$ is written as

$$\boldsymbol{R}(t) = \sum_{\tau=1}^{k} \boldsymbol{B}_\tau \boldsymbol{R}(t-\tau) + \boldsymbol{e}(t) \tag{2.44}$$

where *k* is the number of time-delays (lags) of the autoregression, $\boldsymbol{B}_\tau$ are $n \times n$ matrices of coefficients, and $\boldsymbol{e}(t)$ is the innovation process.

The VAR-LiNGAM model is a combination of both the contemporaneous (lag zero) structural equation model (equation 2.41) and the autoregressive model with time-delays (equation 2.44). The notation is similar to the previous notation: *k* is the number of time-delays (lags) of the autoregression and $\boldsymbol{B}_\tau$ are the $n \times n$ matrices containing the causal effects between returns $\boldsymbol{R}(t-\tau)$ with time lag $\tau = 1, \dots, k$. The $\boldsymbol{B}_\tau$ matrices for $\tau > 0$ correspond to effects from the past to the present, while $\boldsymbol{B}_0$ corresponds to instantaneous effects. The complete VAR-LiNGAM model is

$$\boldsymbol{R}(t) = \sum_{\tau=0}^{k} \boldsymbol{B}_\tau \boldsymbol{R}(t-\tau) + \boldsymbol{e}(t) \tag{2.45}$$

where $\boldsymbol{e}(t)$ are random disturbances. This model is based on three assumptions

1) $\boldsymbol{e}(t)$ are mutually independent and temporally uncorrelated, both with each other and over time.
2) $\boldsymbol{e}(t)$ are non-Gaussian.
3) The matrix $\boldsymbol{B}_0$ corresponds to an acyclic graph.

The model is estimated in two stages. First estimate the following vector autoregressive model using any least squares method

$$\boldsymbol{R}(t) = \sum_{\tau=1}^{k} \boldsymbol{M}_{\tau}\boldsymbol{R}(t-\tau) + \boldsymbol{n}(t) \tag{2.46}$$

where $\boldsymbol{M}_{\tau}$ is the matrix of autoregressive least-squares estimates. Note that $\boldsymbol{M}_{\tau}$ contains no contemporaneous parameter estimates (i.e. where $\tau = 0$). Next, compute the residuals of the model

$$\hat{\boldsymbol{n}}(t) = \boldsymbol{R}(t) - \sum_{\tau=1}^{k} \hat{\boldsymbol{M}}_{\tau}\boldsymbol{R}(t-\tau). \tag{2.47}$$

Then perform a LiNGAM analysis on the residuals to obtain an estimate of the matrix $\boldsymbol{B}_0$, which is the solution to the instantaneous causal model

$$\hat{\boldsymbol{n}}(t) = \boldsymbol{B}_0\hat{\boldsymbol{n}}(t) + \boldsymbol{e}(t). \tag{2.48}$$

Finally, use $\hat{\boldsymbol{B}}_0$ to compute $\boldsymbol{B}_{\tau}$ for $\tau > 0$

$$\hat{\boldsymbol{B}}_{\tau} = \left(\boldsymbol{I} - \hat{\boldsymbol{B}}_0\right)\hat{\boldsymbol{M}}_{\tau} \quad \text{for } \tau > 0, \tag{2.49}$$

and substitute the estimates of these causal effect matrices into equation 2.45.

The VAR-LiNGAM model generates forecasts in the same way that a traditional VAR model generates forecasts. Use historical data and parameter estimates on the right hand side of equation 2.45 to estimate one-step-ahead forecasts.

Portfolios of Independent Factors

The source of variance in the VAR-LiNGAM model is the vector of disturbances $\boldsymbol{e}(t)$ in the model above; each disturbance term can be interpreted as an independent

factor. Any portfolio composed of a set of securities that correspond to the VAR-LiNGAM model can alternatively be thought of as a portfolio composed of a set of independent factors. Identifying and controlling the portfolio's exposure to the independent factors is important because they are the true sources of both risk and return in the VAR-LiNGAM model. As shown in the application section, however, controlling a portfolio's exposure to a particular security does not effectively control the portfolio's exposure to that security's underlying independent factor.

The basic VAR-LiNGAM model definition can be written in terms of the returns on a set of independent factors (Hyvarinen et al. 2010)

$$\widetilde{\boldsymbol{R}}(t) = (\boldsymbol{I} - \boldsymbol{B}_0)\boldsymbol{R}(t) = \sum_{\tau=1}^{k} \boldsymbol{B}_\tau \boldsymbol{R}(t-\tau) + \boldsymbol{e}(t). \qquad (2.50)$$

The independent factor returns $\widetilde{\boldsymbol{R}}(t)$ are composed of a component that contains only information from the past ($\tau > 0$) and a component $\boldsymbol{e}(t)$, which is a vector of independent disturbance terms at $\tau = 0$. When the matrix $\boldsymbol{B}_0$ is interpreted as a directed acyclic graph containing nodes and directed edges, each row of the matrix $(\boldsymbol{I} - \boldsymbol{B}_0)$ subtracts the effects of a node's parents from the effects of the node itself, leaving only the independent factor that drives each node.

With this equation, the relative weights of the underlying independent factors can be identified in a portfolio of correlated securities. This can be done in a similar way that the new factor exposures were found in equation 2.3 by allowing the matrix $(\boldsymbol{I} - \boldsymbol{B}_0)$ to function as the "pick" matrix

$$R_P(t) = \boldsymbol{\omega}^T \boldsymbol{R}(t) = \boldsymbol{\omega}^T (\boldsymbol{I} - \boldsymbol{B}_0)^{-1} (\boldsymbol{I} - \boldsymbol{B}_0) \boldsymbol{R}(t) = \tilde{\boldsymbol{\omega}}^T \tilde{\boldsymbol{R}}(t). \quad (2.51)$$

Thus the relative weights $\tilde{\boldsymbol{\omega}}$ on the independent factors are

$$\tilde{\boldsymbol{\omega}} = (\boldsymbol{I} - \boldsymbol{B}_0)^{-T} \boldsymbol{\omega}. \quad (2.52)$$

These weights can be used in the process of portfolio optimization to control risk as shown below.

Application

In the remainder of this chapter, the VAR-LiNGAM model is used to generate forecasts for the set of the six most widely traded currencies. In addition, estimates from the model are used to find the set of currencies' underlying independent factors. These independent factors are used in the risk contribution framework to provide portfolio optimization constraints. Six portfolios that are constructed using different optimization constraints are compared on a small out of sample data set. The purpose of this brief study is not to find the best performing portfolio but to examine the internal properties of the portfolios that are a result of using different constraints during optimization. The constraints are used to prohibit the allocation of too much wealth to any one risk factor in the portfolio, and some constraints do this more effectively than others. A set of plots shows how effective each constraint set is at controlling each measure of risk contribution. All plots are generated using Gnuplot software (2012).

Description of the Data

Data was obtained from the Sierra Chart FXCM Forex data service using Sierra Chart software (Sierra Chart 2012). This data is based on transactions between the FXCM foreign exchange dealer and its clients. The data set consists of direct quotations with fifteen minute periodicity for the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and U.S. dollar (USD), all in terms of the Swiss franc (CHF). In other words, the exchange rates used in the study are the CHF/X exchange rates where X represents one of the six currencies listed above. These currencies are chosen because they had the largest market turnover rates in 2010 according to the Triennial Central Bank Survey (Bank for International Settlements 2010). Data for the entire year of 2009 (24,916 observations) is used for model estimation while data on 1/4/2010 from 00:00-12:30 (51 observations) is used for a portfolio performance comparison in which forecasts are made and then portfolios are rebalanced with portfolio optimizations based on these forecasts. Missing data is replaced by the most recent observation in each currency series. Log returns are computed by taking the natural logarithm and then first-differencing the exchange rates. The expected values and covariance matrix of the 2009 log returns are shown in table A-1 and table A-2**Error! Reference source not found.**.

VAR-LiNGAM Estimation Results

The two stage estimation technique was used to estimate the VAR-LiNGAM model for the 2009 currency returns. The VAR estimation was performed with SAS,

Version 9.2 (2012). An order 1 VAR was selected using the Hannan-Quinn information criterion and the Schwarz's Bayesian criterion. Both criterions were negligibly better for order 2, but order 1 was chosen for the sake of parsimony. The LiNGAM procedure was then performed on the VAR residuals using the MATLAB code provided by Shimizu et al. (2006). The same code was used to prune the $\boldsymbol{B}_0$ matrix at alpha level 0.01 using the Bonferroni correction technique (Shaffer 1995). The expected returns for both the VAR residuals and the independent factors are presented in table A-1 while their respective covariance matrices are shown in table A-3 and table A-4. The pruned $\boldsymbol{B}_0$ matrix is shown in table A-**Error! Reference source not found.**5; this matrix corresponds to the directed acyclic graph shown in figure A-1. The $\boldsymbol{B}_1$ matrix is shown in table A-6 and the autoregressive matrix $\boldsymbol{M}_1$ is shown in table A-7. This procedure was repeated for the log returns in the entire year of 2010, and the resulting directed acyclic graph is shown in figure A-2 for comparison. The directed acyclic graphs in figures A-1 and A-2 show the instantaneous causal effects between the currencies and were generated using Graphviz software (2012). Edges of the graphs are directed to represent causal flow. For instance, an edge $X \rightarrow Y$ indicates that a movement of currency X causes a movement of currency Y.

As evidence that the VAR-LiNGAM non-Gaussian assumption holds, a Kolmogorov-Smirnov test performed on each currency's corresponding independent factor confirmed that the null hypothesis of normality was rejected with p-value less than 0.01 for each factor.

Portfolio Optimization with Constraints

A variety of different constraint sets can be added to the canonical mean-variance model to achieve varying portfolio objectives. The purpose of the constraints described here is to increase portfolio diversification so that the portfolio's risk is not concentrated too greatly in any one risk factor. In each case, the metric or attribute being constrained characterizes the diversification differently. Constraints on the following attributes are added to the canonical mean-variance model (equations 2.11-2.12): relative portfolio weights, independent factor relative portfolio weights, percentage marginal risk contributions, independent factor percentage marginal risk contributions, and percentage total risk contributions. All optimizations are performed using AMPL (2011).

The relative portfolio weight constraints and the independent factor relative weight constraints are created by normalizing the relative weights with the sum of the risky asset weights (excluding the risk free asset). This normalization attempts to make the portion of the portfolio invested in risky assets more diversified. The limits on the relative weights are chosen to be 0.25 so that no more than 25% of the risky portion of the portfolio is invested (long or short) in any one risky asset. This leads to the addition of the following constraint set to the canonical mean-variance model

$$-0.25 \leq \frac{b_i}{\sum_{i=1}^{6} b_i} \leq 0.25 \quad for\ i = 1, \dots, 6 \tag{2.53}$$

where $\boldsymbol{b}$ is defined as either $\boldsymbol{\omega}$ or $\tilde{\boldsymbol{\omega}}$. This constraint set is referred to as the relative portfolio weight constraint set (WC) when $\boldsymbol{b}$ is defined as $\boldsymbol{\omega}$ and the independent factor relative weight constraint set (WIC) when $\boldsymbol{b}$ is defined as $\tilde{\boldsymbol{\omega}}$.

Limits placed on each factor's percentage marginal risk contribution in a portfolio optimization prohibit any one factor's marginal risk contribution from dominating the portfolio. The limits on the percentage marginal risk contributions are chosen to be 0.25 to be consistent with the constraints on the relative portfolio weights above. The resulting constraint set is

$$-0.25 \leq \frac{b_i * \frac{\delta \mathcal{R}(\boldsymbol{b})}{\delta b_i}}{\mathcal{R}(\boldsymbol{b})} \leq 0.25 \quad for\ i = 1, \dots, 6 \tag{2.54}$$

where $\boldsymbol{b}$ is defined as either $\boldsymbol{\omega}$ or $\tilde{\boldsymbol{\omega}}$. For this application, the portfolio's standard deviation is chosen as the measure of risk (equation 2.21). This constraint set is referred to as the marginal risk contribution constraint set (MRC) when $\boldsymbol{b}$ is defined as $\boldsymbol{\omega}$ and the independent factor marginal risk contribution constraint set (MRIC) when $\boldsymbol{b}$ is defined as $\tilde{\boldsymbol{\omega}}$.

Limits placed on each independent factor's percentage total risk contribution in a portfolio optimization prohibit any one independent factor's total risk contribution from dominating the portfolio. The total risk contributions used here are the components of the total risk decomposition of the portfolio variance (equation 2.25). The limits on the percentage total risk contributions are chosen to be 0.25 to be consistent with the previously defined constraints. The resulting constraint set is

$$\frac{\mathcal{R}(\tilde{\omega}_i)}{\mathcal{R}(\tilde{\boldsymbol{\omega}})} \leq 0.25 \quad for\ i = 1, \dots, 6. \tag{2.55}$$

This constraint set is referred to as the independent factor total risk contribution constraint set (TRC).

Efficient Frontiers

In this section, an efficient frontier is constructed using the unconstrained canonical mean-variance model and using each of the constraints discussed above appended to the canonical mean-variance model one at a time. To simplify the computation of the efficient frontiers, the risk free rate is set equal to zero and all estimates are computed using data from the entire year of 2009. The portfolio's expected return is computed using the expected value of the currency log returns

$$R_P = \boldsymbol{\omega}^T E\{\boldsymbol{R}(t)\} \quad t \in 2009, \tag{2.56}$$

and the variance of the portfolio's return is computed using the covariance matrix of the currency log returns

$$\sigma_P^2 = \boldsymbol{\omega}^T Cov\{\boldsymbol{R}(t)\}\boldsymbol{\omega} \quad t \in 2009. \tag{2.57}$$

The matrix $\boldsymbol{B}_0$, which is required to compute the relative weights on the independent factors $\tilde{\boldsymbol{\omega}}$ via $(\boldsymbol{I} - \boldsymbol{B}_0)^{-T}$, was also estimated using data from the entire year of 2009**.** In each of the optimizations the efficient frontier is constructed by varying λ across the interval [0, 1]. The resulting frontiers reflect the year 2009 in general and no particular point in time.

The set of efficient frontiers is shown in figure 2.1. The frontiers presented here are linear because a risk free asset is present in the portfolio. In comparison, when a risk free asset is not present, an efficient frontier is the upper edge of a hyperbola shaped region that surrounds all possible portfolios. The efficient frontier lines in figure 2.1 start at the risk free rate on the y-axis and continue through a tangency portfolio on the upper edge of their respective hyperbola shaped regions (not shown) that surround the set of all

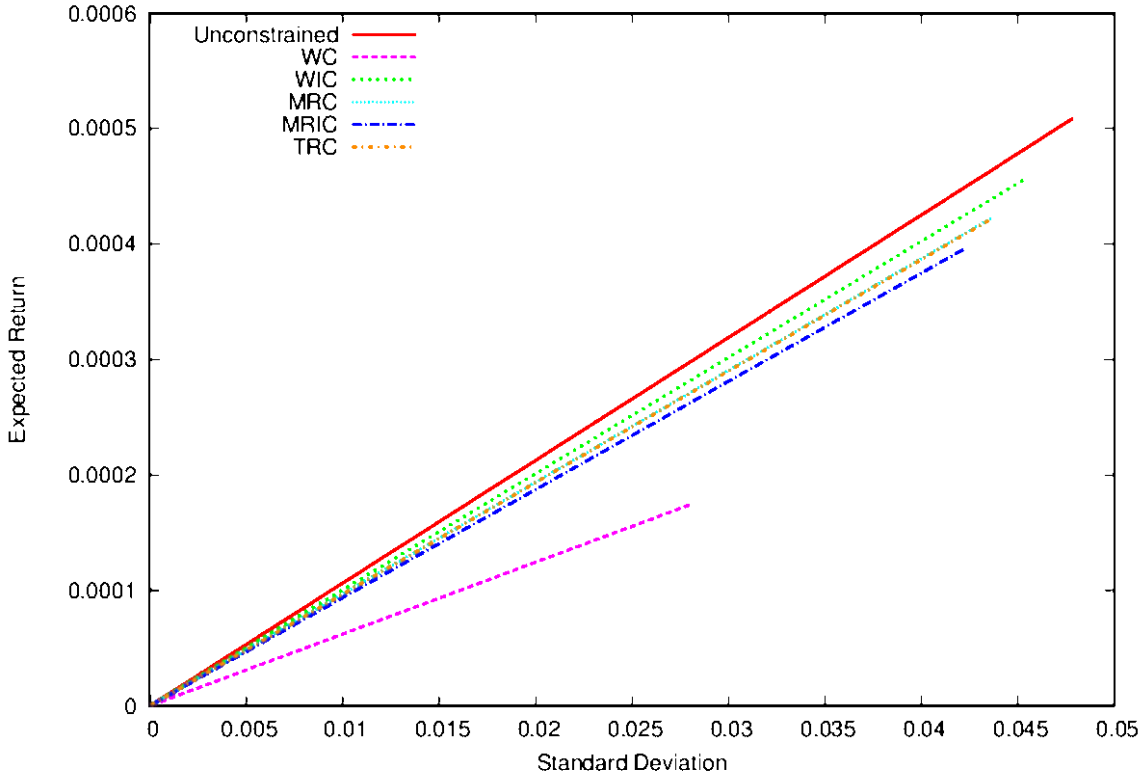


Figure 2.1. Comparison of efficient frontiers
Note: The frontiers are generated using constrained mean-variance optimization with data from 2009. Each frontier in the plot is labeled according to the constraint set used in its creation. The constraint sets and labels are as follows: the relative portfolio weight constraint set (WC), the independent factor relative weight constraint set (WIC), the marginal risk contribution constraint set (MRC), the independent factor marginal risk contribution constraint set (MRIC), and the independent factor total risk contribution constraint set (TRC).

possible risky portfolios. Unconstrained frontiers are ubiquitous in finance, and frontiers created with relative portfolio weight constraints are considered occasionally. Frontiers created with marginal risk contribution constraints are rarely used, and the frontiers created with independent factor constraints are unique to this study.

There is a cost in terms of efficiency of adding constraints to the optimization. The unconstrained frontier is the most efficient, i.e. has the greatest portfolio expected return per standard deviation of portfolio return, followed by (in order) the frontiers generated with the following constraint sets: the independent factor relative weight constraint set (WIC), the marginal risk contribution constraint set (MRC), the independent factor total risk contribution constraint set (TRC), the marginal risk contribution of independent factor constraint set (MRIC), and the relative portfolio weight constraint set (WC). Note that the MRC and TRC constrained frontiers overlap.

All but the relative portfolio weight constraint set produce frontiers that are near to one another and relatively near to the unconstrained frontier. The relative portfolio weight constrained frontier's large distance from the unconstrained frontier indicates that the cost of this constraint set is much greater than the cost of the other constraint sets.

Portfolio Performance

The performance of each of these portfolios is compared using data that spans 00:00-12:30 on 1/4/2010; the realized returns for each of the six currencies during this time period are shown in figure A-3. It is assumed that the risk free rate is zero and that there are no transaction costs. The VAR-LiNGAM estimation results from 2009 of $E\{\boldsymbol{n}(t)\}$, $Cov\{\boldsymbol{n}(t)\}$, and $\boldsymbol{M}_1$ are used in the time series model that serves as the base for the portfolio optimizations. The 2009 estimate for $\boldsymbol{B}_0$ is used to compute the relative weights on the independent factors $\tilde{\boldsymbol{\omega}}$. For each period's portfolio optimization, the most

recently observed return $\boldsymbol{R}(t)$ is used to compute the next period's conditional expected value $E\{\boldsymbol{R}(t+1) \mid \boldsymbol{M}_1, \boldsymbol{R}(t)\}$. This computation is as follows

$$E\{\boldsymbol{R}(t+1) \mid \boldsymbol{M}_1, \boldsymbol{R}(t)\} = E\{\boldsymbol{M}_1\boldsymbol{R}(t) + \boldsymbol{n}(t) \mid \boldsymbol{M}_1, \boldsymbol{R}(t)\} \qquad (2.58)$$

$$= \boldsymbol{M}_1\boldsymbol{R}(t) + E\{\boldsymbol{n}(t)\}$$

The 2009 covariance matrix of the residuals is used for each portfolio optimization and is assumed to remain constant throughout 1/4/2010.

Each portfolio is created with $\lambda = 0.99$ in its corresponding optimization model for every time period in the simulation; this high degree of risk aversion was chosen so that the portfolios do not become too volatile. Each portfolio starts with an initial wealth in cash (unallocated to other currencies) of 100 CHF. A comparison of the portfolio performance is shown in figure 2.2. The cumulative wealth of each portfolio at any particular point in time in figure 2.2 can be explained by comparing the portfolio holdings shown in figure A-4 with the realized returns shown in figure A-3. A portfolio experiences a large change in value when it has a large position in a currency that has a large realized return. For example, in period 30 the unconstrained portfolio had a substantial loss in value (see figure 2.2) because it had a large short (i.e. negative) position in the euro (see figure A-4) and the euro had a large positive return (see figure A-3). The value of the other portfolios did not change much in period 30 because none of them had any substantial positions.

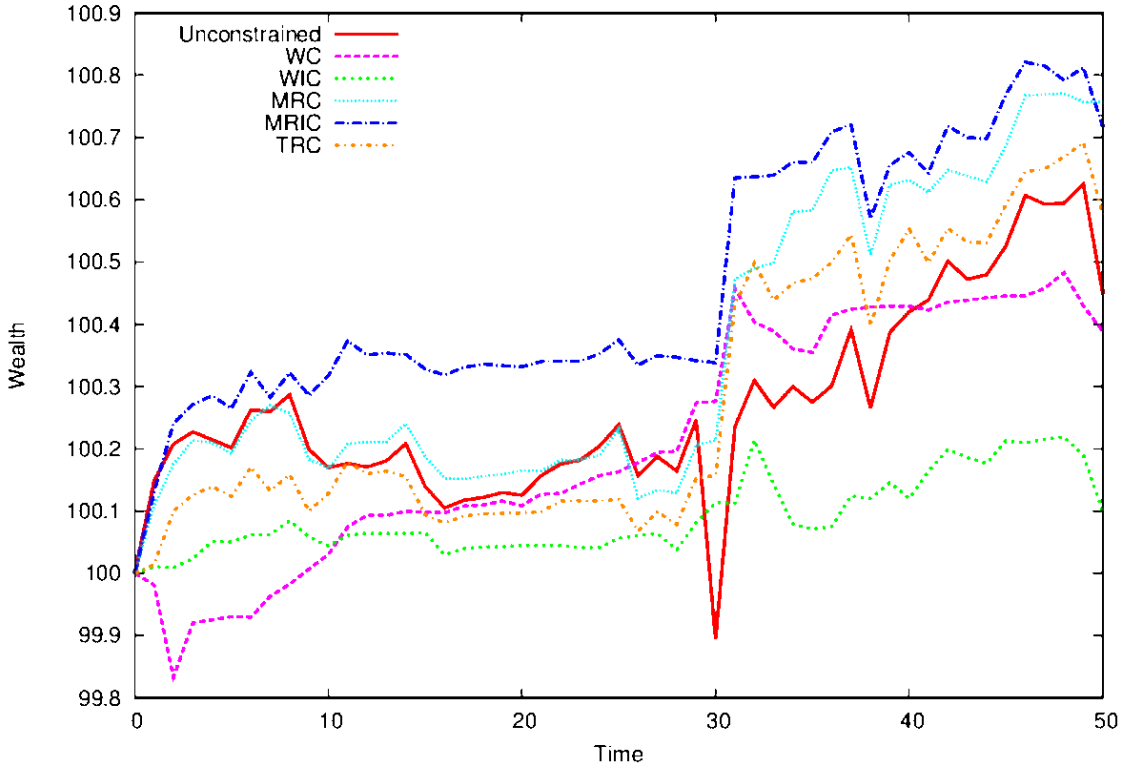


Figure 2.2. Portfolio performance comparison on 1/4/2010 from 00:00-12:30
Note: For every point in time in the plot, the VAR-LiNGAM model is used to generate forecasts, and then constrained mean-variance optimization is used to rebalance each portfolio. Initial portfolio wealth is 100 CHF. Each portfolio in the plot is labeled according to the constraint set used in its creation. The constraint sets and labels are as follows: the relative portfolio weight constraint set (WC), the independent factor relative weight constraint set (WIC), the marginal risk contribution constraint set (MRC), the independent factor marginal risk contribution constraint set (MRIC), and the independent factor total risk contribution constraint set (TRC).

In period 31, the unconstrained portfolio made up the lost ground from the previous period by having an even larger short position in the euro and a large short position in the yen; the yen's negative return in period 31 was large, while the euro's was modest. The WC, MRC, MRIC, and TRC portfolios also experienced large gains in period 31 (see figure 2.2) by having large short positions in the yen and euro. In period

50, all portfolios lost value because they had long positions in the pound sterling, which had a large negative return. Overall, on this small data set none of the portfolios performed exceptionally better than the others, but the portfolio generated with the independent factor relative weight constraint set (WIC) did seem to lag behind.

Each of figures A-4 through A-8 contains six plots, and each plot corresponds to a portfolio constructed with one of constraints discussed above. All of the plots within a figure display the same attribute (e.g. total risk contribution in figure A-8) for the currency positions contained within a portfolio.

Plots of portfolio currency allocations are shown in figure A-4; each line in the figure represents the total value of a particular currency holding denominated in Swiss francs. At some points in time and for all constraints except the weight constraints, the allocation to a specific currency is disproportionately large. Only with the relative weight constraints are the allocations of all currencies kept within a narrow boundary.

Independent factor relative portfolio weights over time are displayed in figure A-5. The independent factor relative weight constraint set (WIC) and the relative portfolio weight constraint set (WC) appear to be the most effective at keeping the independent factor relative portfolio weights constrained. The other constraint sets allow some of the independent factor relative portfolio weights to become many times larger than the others; this concentration of wealth in a small number of independent factors is undesirable because it represents a loss of diversification.

Percentage marginal risk contribution plots are shown in figure A-6, and independent factor percentage marginal risk contribution plots are shown in figure A-7.

None of the plots in either of these figures indicate that the marginal risk contributions become as divergent as some of the independent factor relative portfolio weights do in figure A-5.

Independent factor total risk contribution plots are displayed figure A-8. In addition to the independent factor total risk contribution constraint set (TRC), the independent factor relative weight constraint (WIC) and marginal risk contribution of independent factor constraint (MRIC) sets do a moderate job of controlling total risk contributions. The total risk contributions that result from optimizations using other constraint sets are at times out of proportion. For example, in the weight constrained (WC) portfolio, the euro and U.S. dollar independent factors contribute most to the portfolio's total risk. This is an indication that while weight constraints are one of the most popular ways to control portfolio exposures, they can be very ineffective at controlling risk contributions.

Conclusions

With traditional portfolio analysis, a portfolio manager's life is complicated by the fact that risk can only be analyzed relative to an existing portfolio. Once a portfolio is somehow constructed, marginal risk analysis is confusing because changing the size of one position not only changes its marginal risk contribution but changes the marginal risk contributions from the other positions as well. To further complicate matters, some positions have negative marginal risk contributions so that increasing such a position's size will actually reduce the risk of the portfolio at the margin.

Total risk contributions, as introduced in this chapter, are simpler to compute and easier to interpret than marginal risk contributions. Each total risk contribution is non-negative and can be interpreted as having a potentially adverse effect on the portfolio. Furthermore, in the total risk contribution framework, each position contributes independently to the risk of the total portfolio so that changing the size of one position does not affect the risk contributions of the other positions. This allows a portfolio manager to know exactly how much risk each position (in an independent factor) will contribute to the portfolio. In the example application, percentage total risk contributions are constrained to be less than or equal to 25% of the total portfolio's risk. This is equivalent to a portfolio manager deciding that no more than 25% of the portfolio's risk should be allocated to any one independent risk factor. This decision takes place before the portfolio is created, and the percentage risk allocations are not relative to an existing portfolio as they would be when using the traditional marginal risk contribution framework.

When a set of securities conforms closely to the assumptions of independent component analysis, the independent factors can be thought of as the true sources of risk and return. In this case, the total risk contribution framework is recommended for use in both portfolio construction (e.g. optimization) and portfolio analysis because of its simplicity and conceptual clarity.

In the application section of this chapter, the market structure underlying a set of currencies is found, and from this structure the set of independent factors is identified. The total risk contributions from the total risk decomposition of a portfolio's variance

are computed using this set of independent factors. Portfolios are constructed on sample data sets by constraining the relative portfolio weights of the independent factors and the total risk contributions during mean-variance optimizations. The efficient frontiers and the properties of these portfolios as they change over time compare favorably to those of other portfolios generated by more conventional constraints during mean-variance optimization. A comparison of the internal properties of these portfolios shows that, in this case, the only way to effectively control independent risk contributions is to put constraints on the independent risk contributions themselves.

CHAPTER III

FOREIGN EXCHANGE MARKET STRUCTURE

Introduction

The objective of this chapter is to discover the causal relationships between the six most widely traded currencies and to observe how these causal relationships change over time. The PC, Fast Causal Inference (FCI), and LiNGAM causal search algorithms are used to find causal relationships, and the results are presented graphically. A secondary goal of the study is to determine whether any latent common causes influence the foreign exchange market; this is accomplished with the FCI algorithm.

Little is known about the foreign exchange market in spite of it being the focus of an abundant amount of research. Previous literature finds that exchange rates are disconnected from macroeconomic fundamentals, for instance see Meese and Rogoff (1983), Frankel and Rose (1994), and Cheung, Chinn, and Pascual (2005). In other words, exchange rate models based on macroeconomic fundamentals have no explanatory or predictive power. Cheung, Chinn, and Pascual (2005) examines several exchange rate models at a variety of forecast horizons and concludes that no model consistently outperforms a random walk.

Evans (2012) theorizes that the dominant drivers of changes in real exchange rates are latent variables, which he names “dark matter”. These latent variables have an impact on real exchange rates via expected excess returns or the expected long-run real exchange rates. The model used by Evans (2012) reveals that the only observable

variables to have even a marginal effect on real exchange rates for major currencies are real interest rate differentials.

In a different strand of literature, authors use cointegration analysis to find relationships between currencies. These studies produce a wide range of results; some find cointegration in the foreign exchange market, some do not, and others have mixed results. For example, Norrbin (1996) finds that most currencies in the European Monetary System are cointegrated during the period 1979-1992. Rapp and Sharma (1999) finds that none of the G-7 currencies are cointegrated during the period 1973-1996, whereas Phengpis (2006) finds mixed and unstable results during the 1992–1993 European and the 1997–1998 Asian currency crises. For more references to cointegration literature that examines the foreign exchange market before introduction of the euro, see Kuhl (2010).

Kuhl (2010) is the first to search for cointegration in the foreign exchange market since the formation of the European Monetary Union. Using data for the Australian dollar, euro, Great Britain pound sterling, Japanese yen, and Swiss franc all in terms of the United States dollar during the period from January 1999 to June 2007 (after the introduction of the euro), the study finds that the euro-US dollar exchange rate is weakly exogenous and cointegrates with the Australian dollar-US dollar and with the British pound–US dollar. Hence, the study asserts that causality runs from the euro-US dollar to the other two exchange rates.

Related Literature

In contrast to the previous literature on foreign exchange, the analysis contained within this paper uses graph-theoretic search algorithms to uncover the relationships amongst a set of currencies. These techniques originated in the computer science literature and have been developed into a framework for finding causal relationships amongst a set of variables using only observational data. A thorough introduction to these ideas can be found in Pearl (2000).

In most studies that apply causal search methods to economic data, the innovations from vector autoregressive (VAR) models or vector error correction (VEC) models are used to find directed acyclic graphs (DAGs) that represent the contemporaneous causal relationships between selected economic variables. The PC algorithm (explained below) is the most popular technique for finding these causal relationships.

Some examples of this literature are as follows. Akleman, Bessler, and Burton (1999) uses the PC algorithm to find DAGs that represent the causal relationships of exchange rates on U.S. corn exports. Haigh and Bessler (2004) finds that both the PC algorithm and a graph scoring technique agree on the same contemporaneous causal structure underlying soybean prices and barge transportation costs. Kim, Leatham, and Bessler (2007) looks explicitly for causal structural change across time and finds that the contemporaneous causal structure of a model is different before and after a structural break. Moneta (2008) modifies the PC algorithm in an attempt to provide more stability

to the algorithm's edge orientation procedure and applies this modified algorithm to US economic data.

The remaining part of this paper is divided into two sections. In the first section, directed graphs and their connection to causal modeling is explained. Three causal search algorithms are then presented and their usage with time series models is discussed. The second section of the paper contains an application of these techniques to a set of the six most widely traded currencies. Each of the years 2009, 2010, and 2011 are modeled separately and the results are compared.

Directed Acyclic Graphs

A directed graph is a diagram that depicts a set of vertices (nodes) V connected by a set of directed edges (links) E with each edge connecting a pair of vertices, e.g. for $\{\mathrm{A}, \mathrm{B}\} \in \mathrm{V}$, $A \rightarrow B$. In a directed acyclic graph (DAG), there is no path (sequence of directed edges) that leads from a vertex back to itself. In other words, a DAG contains no directed cycles, e.g. $A \rightarrow B \rightarrow A$.

Directed acyclic graphs are often used to represent the causal relationships among a set of random variables. In the causal setting, each vertex represents a variable, and all directed edges are drawn with arrowheads that indicate the direction of causal flow, e.g. $A \rightarrow B$ indicates that variable A causes variable B. A set of variables V is causally sufficient if every common cause of two variables in V is also in the set V. A DAG G on a causally sufficient set of variables V is a causal DAG if and only if there is

an edge from A to B in G if and only if A is a direct cause of B relative to V (Spirtes, Glymour, and Scheines 2001).

Several algorithms have been developed to find the causal structure underlying a set of variables. These algorithms use probability distributions (specifically, conditional independence information) on a set of variables to construct a graphical model illustrating the variables' causal relationships. Usually the output is one DAG that does not uniquely identify the set of causal relationships. The best that can be done using observed conditional independence information is to find a set of DAG structures that are observationally equivalent. Sometimes an equivalence class of DAGs is represented by a partially directed DAG in which some edges are not directed. The directed edges are common to all members of the equivalence class while the undirected edges are directed one way in some members of the equivalence class and the opposite way in other members (Pearl 2000).

In the presence of latent variables or selection bias, a partial ancestral graph (PAG) may be used to represent partial causal information that is shared by a class of DAGs. PAGs contain conditional independence information and partial information about ancestor relationships entailed by a class of DAGs.

PC Algorithm

There are a number of causal search methods for recovering the graphical structure underlying a set of variables. The PC algorithm, named after its authors Peter and Clark (Spirtes, Glymour, and Scheines 2001), is a classic search method used in

much of the previous literature. The PC algorithm assumes that the true causal graph is acyclic and that there are no hidden common causes of any two variables. The Fisher's z test that is commonly used in implementations of the PC algorithm assumes that each variable has a Normal distribution and that direct causal influences between variables are linear.

The PC algorithm is implemented in two stages. In the first stage, the algorithm starts with a complete (fully connected) undirected graph. Then, edges are removed with a sequence of conditional independence tests. For every pair of variables, the algorithm searches for a set of the other variables that renders the pair conditionally independent. If such a set is found then the edge between the pair of variables is removed from the graph. During this edge removal procedure some undirected edges are directed as a precursor to the second stage of the algorithm. Fisher's z is commonly used instead of a test for conditional independence; this tests whether conditional correlations are different from zero. Fisher's z equation is

$$z(\rho(i,j|k),n) = \left[\frac{1}{2}\sqrt{n-|k|-3}\right] ln\left\{\frac{|1+\rho(i,j|k)|}{1-\rho(i,j|k)}\right\} \qquad (3.1)$$

where $\rho(i,j|k)$ is the population correlation between *i* and *j*, *n* is the number of observations, and $|k|$ is the number of variables in set *k*. If *i*, *j*, and *k* are normally distributed and $r(i,j|k)$ is the sample conditional correlation of *i* and *j* given *k* then $z(\rho(i,j|k),n) - z(r(i,j|k),n)$ has a standard normal distribution.

In the second stage of the algorithm, the resulting partially directed graph of the first stage is then directed more fully based on a set of rules. This set of rules is presented in Meek (1995) and Pearl (2000).

The output of the PC algorithm is a graph that may contain three types of edges. A directed edge between a pair of variables indicates that one variable is a direct cause of the other variable. An undirected edge between a pair of variables indicates that the PC algorithm cannot determine which variable causes the other. A doubly directed edge between a pair of variables indicates that one of the algorithm's assumptions has not been met or that some statistical decisions are inconsistent because the sample size is not large enough. The PC algorithm is implemented in the TETRAD IV software package (Glymour et al. 2012).

Fast Causal Inference (FCI) Algorithm

The FCI algorithm of Spirtes, Glymour, and Scheines (2001) is similar to the PC algorithm; it assumes that the true causal graph is acyclic and that there might be hidden (latent) variables or selection bias. A partial ancestral graph (PAG) is produced by the algorithm using the causal Markov assumption, the causal faithfulness assumption, and the population inference assumption. A PAG represents the ancestor relationships among a set of variables that are shared by a set of observationally equivalent DAGs. Variable A is an ancestor of variable B when there is a directed path from A to B.

An edge in a PAG can have three kinds of endpoints: "∘", "-", or ">". An edge between two measured variables indicates that the variables are not independent

conditional on any set of measured variables. The following four types of edges can be made by combining the different types of endpoints: A ∘ – ∘ B, A ∘→ B, A → B, or A ↔ B. Assuming that there is no selection bias, the interpretation of edge types may be summarized as follows (let "*" stand for any of the three types of endpoints); for more a more detailed description, see Spirtes, Glymour, and Scheines (2001)

1. An edge between A and B is oriented as A → B only if A is an ancestor of B.
2. An edge between A and B is oriented as A ∗→ B only if B is not an ancestor of A.
3. An "∘" on the end of an edge places no restriction on the possible ancestor relationships.

Note that an edge does not necessarily mean that one variable directly causes another variable. The existence of an edge implies that the causal pathway may contain latent variables. In the absence of selection bias, an edge A ↔ B implies that A and B possibly share a common cause.

Like the PC algorithm the FCI algorithm is implemented in two stages. In the first stage, edges are removed, and in the second stage, edges are directed. The particular implementation of the FCI algorithm used here employs Fisher's z as the test for conditional independence. The FCI algorithm is implemented in the TETRAD IV software package (Glymour et al. 2012).

LiNGAM Algorithm

The LiGAM algorithm assumes that the observed variables can be arranged in a causal order so that the data generating process can be represented by a DAG, that the value assigned to each variable is a linear function of values assigned to variables positioned earlier in the causal order, that there are no latent common causes, and that the disturbance terms are mutually independent with non-Gaussian distributions and non-zero variances (Shimizu et al. 2006). The non-Gaussian assumption is important because this allows LiNGAM to estimate the full causal model with no undetermined parameters.

LiNGAM assumes that the observed variables $\boldsymbol{x}$ are linear functions of the disturbance variables $\boldsymbol{e}$. When the mean is subtracted from each variable, this is expressed as

$$\boldsymbol{x} = \boldsymbol{B}\boldsymbol{x} + \boldsymbol{e}. \tag{3.2}$$

Solving for $\boldsymbol{x}$, this becomes

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{e} \tag{3.3}$$

where $\boldsymbol{A} = (\boldsymbol{I} - \boldsymbol{B})^{-1}$. Equation 3.2 in addition to the assumption that the disturbance terms are independent and have non-Gaussian distributions is the independent component analysis (ICA) model. The ICA model has the following two indeterminancies that must be resolved before a graphical model can be constructed: neither the order nor the scaling of the independent components is defined. LiNGAM resolves both of these issues by permuting and normalizing the ICA mixing matrix to

obtain a matrix $\boldsymbol{B}$ containing the DAG connection strengths. The graphical representation of this matrix is the causal DAG model.

Because LiNGAM uses the non-Gaussian information contained in the disturbance terms, its output is just one DAG instead of an equivalence class of DAGs. As noted earlier, this output includes parameter estimates for the linear model. The LiNGAM procedure is implemented in MATLAB code provided by Shimizu et al. (2006).

VAR Models

The learning algorithms described above are not designed to be used with time series data. The algorithms are commonly used in conjunction with a vector autoregression (VAR) model so that the VAR model accounts for the relationships between variables across time periods and the learning algorithm finds the contemporaneous causal structure within the most recent period of time. The marriage of these two methods is carried out in practice by first fitting a VAR model to the data and then using its innovations as input into a learning algorithm.

An autoregressive time series model is typically built using a time series of return observations, which is represented in vector form as

$$\boldsymbol{R}(t) = \begin{pmatrix} R_1(t) \\ \vdots \\ R_N(t) \end{pmatrix}, \qquad t \in 1, \dots, T. \tag{3.4}$$

where $R_i(t)$ is the return on a particular asset $i \in 1, \dots, N$ at time $t \in 1, \dots, T$. Using this notation, an autoregressive model of multivariate time series return data is written as

$$\boldsymbol{R}(t) = \sum_{\tau=1}^{k} \boldsymbol{B}_{\tau}\boldsymbol{R}(t-\tau) + \boldsymbol{e}(t) \tag{3.5}$$

where k is the number of time-delays (lags) of the autoregression, $\boldsymbol{B}_{\tau}$ are $n \times n$ matrices of coefficients, and $\boldsymbol{e}(t)$ is the innovation process.

To find an estimate $\widehat{\boldsymbol{n}}(t)$ of the innovation process, estimate the following vector autoregressive model using any least squares method

$$\boldsymbol{R}(t) = \sum_{\tau=1}^{k} \boldsymbol{M}_{\tau}\boldsymbol{R}(t-\tau) + \boldsymbol{n}(t) \tag{3.6}$$

where $\boldsymbol{M}_{\tau}$ is the matrix of autoregressive least-squares estimates. Note that $\boldsymbol{M}_{\tau}$ contains no contemporaneous parameter estimates (i.e. where $\tau = 0$). An estimate of the innovation process can now be computed as

$$\widehat{\boldsymbol{n}}(t) = \boldsymbol{R}(t) - \sum_{\tau=1}^{k} \boldsymbol{M}_{\tau}\boldsymbol{R}(t-\tau). \tag{3.7}$$

This estimate is then used as input to a structural learning algorithm to find the contemporaneous structure underlying the original vector of returns.

VAR-LiNGAM

LiNGAM can be combined with the VAR model in a specific way to so that the VAR model becomes fully identified as described in Hyvarinen et al. (2010); in the following text, this combined model is called VAR-LiNGAM. The VAR-LiNGAM model is a combination of the autoregressive model with time-delays (equation 3.5) and

the following structural equation model, which does not consider the time series structure in data

$$\boldsymbol{R} = \boldsymbol{B}\boldsymbol{R} + \boldsymbol{e} \tag{3.8}$$

where $\boldsymbol{e}$ is a vector of disturbances and the diagonal of $\boldsymbol{B}$ is defined to be zero. The complete VAR-LiNGAM model is

$$\boldsymbol{R}(t) = \sum_{\tau=0}^{k} \boldsymbol{B}_{\tau}\boldsymbol{R}(t-\tau) + \boldsymbol{e}(t) \tag{3.9}$$

where $\boldsymbol{e}(t)$ are random disturbances. The notation is similar to the previous notation: k is the number of time-delays (lags) of the autoregression and $\boldsymbol{B}_{\tau}$ are the $n \times n$ matrices containing the causal effects between returns $\boldsymbol{R}(t-\tau)$ with time lag $\tau = 0, \ldots, k$. The $\boldsymbol{B}_{\tau}$ matrices for $\tau > 0$ correspond to effects from the past to the present, while $\boldsymbol{B}_0$ corresponds to instantaneous effects.

The VAR-LiNGAM model is based on three assumptions

1) $\boldsymbol{e}(t)$ are mutually independent and temporally uncorrelated, both with each other and over time.
2) $\boldsymbol{e}(t)$ are non-Gaussian.
3) The matrix $\boldsymbol{B}_0$ corresponds to an acyclic graph.

The model is estimated in two stages. First estimate the vector autoregressive model and compute the residuals of the model as described above. Then perform a LiNGAM analysis on the estimate of the innovation process to obtain an estimate of the matrix $\boldsymbol{B}_0$, which is the solution to the instantaneous causal model

$$\hat{\boldsymbol{n}}(t) = \boldsymbol{B}_0\hat{\boldsymbol{n}}(t) + \boldsymbol{e}(t). \tag{3.10}$$

Finally, use $\boldsymbol{B}_0$ to compute $\boldsymbol{B}_\tau$ for $\tau > 0$

$$\widehat{\boldsymbol{B}}_\tau = \left(\boldsymbol{I} - \widehat{\boldsymbol{B}}_0\right)\widehat{\boldsymbol{M}}_\tau \quad \text{for } \tau > 0. \tag{3.11}$$

Application

In the remainder of the paper, the three learning algorithms are applied to a set of foreign exchange futures contract data. The causal structure underlying the futures contracts for the six most widely traded currencies is found for each of the years 2009, 2010, and 2011. The graphical structures found by each algorithm are compared, and the structure of the foreign exchange futures market as it changes over time is illustrated.

Description of the Data

Data is obtained from the Sierra Chart historical intraday futures data service using Sierra Chart software (Sierra Chart Software 2012). The data set consists of futures contracts that are traded on the CME Group exchange for the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the Swiss franc (CHF). The March, June, September, and December futures contracts for the years 2009-2011 in addition to the 2012 March contract of each currency are in the dataset. The original data has one minute periodicity and is aggregated across time into fifteen minute intervals so that the resulting data used in this analysis has fifteen minute periodicity.

Sierra Chart software was used to join each currency's future contracts into a single continuous time series for the corresponding currency; for instance, all futures contracts for the AUD (March 2009, June 2009, ..., March 2012) were joined in sequence to form a single continuous time series for the AUD. The resulting six time series were then all converted to be in terms of the Swiss franc (CHF). Thus, the data set used for the analysis consists of observations from 2009-2011 of the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the United States dollar (USD) all in terms of the Swiss franc (CHF).

These currencies are chosen because they had the largest market turnover rates in 2010 according to the Triennial Central Bank Survey (Bank for International Settlements 2010). Application of the causal learning algorithms is performed separately on the data in each of the years 2009, 2010, and 2011. Missing data is replaced by the most recent observation in each currency series. Log returns are then computed by taking the natural logarithm and first-differencing the exchange rates (in that order). All log returns in all time periods are stationary based on Dickey Fuller tests. The expected values and covariance matrix of the log returns for each year are shown in table B-1 and table B-2.

Estimation Results

A different VAR model was estimated for each of the years 2009, 2010, and 2011 using SAS, Version 9.2 (2012). The order of each VAR model was chosen to be order 1 by using the Hannan-Quinn information criterion and the Schwarz's Bayesian

criterion. Both criterions were negligibly better for order 2; order 1 VAR models, however, are used for the sake of parsimony.

Estimates of the innovation processes were then computed as described in equation 3.7. The three structural learning algorithms were then applied to each year's estimated innovation process. This results in three postulated graphical structures (one for each year) for each structural learning algorithm with a total of nine graphical structures altogether. The graphical structures for the years 2009, 2010, and 2011 are presented in figure 3.1, figure 3.2 and figure 3.3. In all searches, the PC and FCI algorithms were set to use an alpha level of 0.001.

As evidence that the VAR-LiNGAM non-Gaussian assumption holds, a Kolmogorov-Smirnov test performed on each currency's corresponding independent factor confirmed that the null hypothesis of normality was rejected with p-value less than 0.01 for each factor. It is noteworthy that the PC and FCI algorithms performed well without using non-Gaussian information; they found graphical skeleton structures that were very similar to those found by LiNGAM as can be seen in figures 3.1-3.3.

The expected returns for the VAR residuals are presented in table B-3 and the covariance matrices are shown in table B-4.

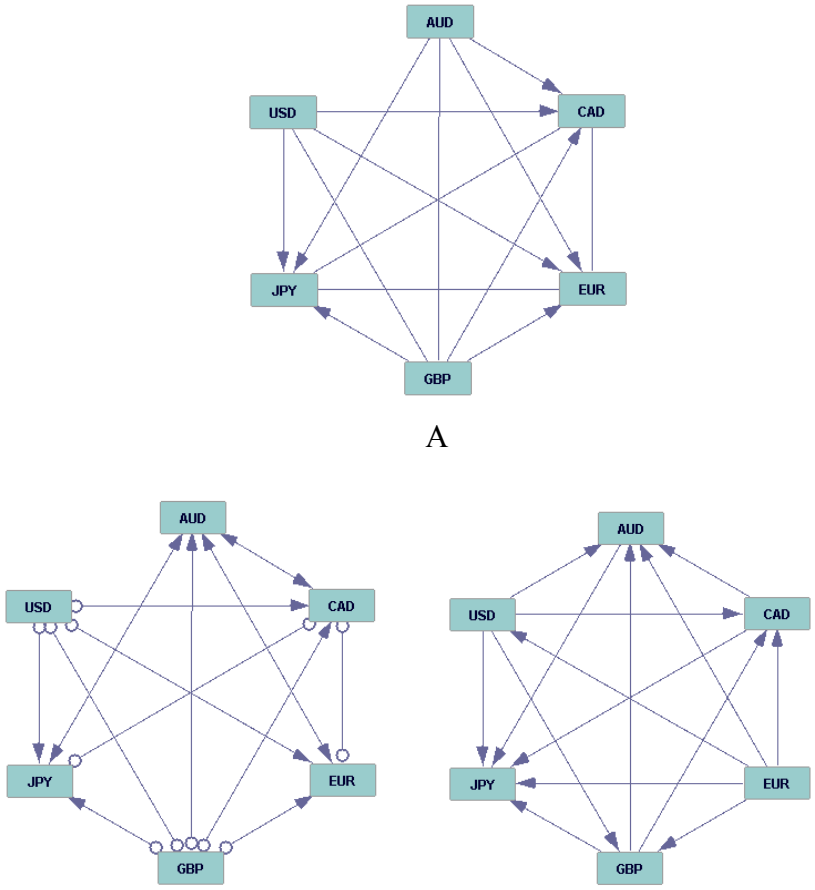
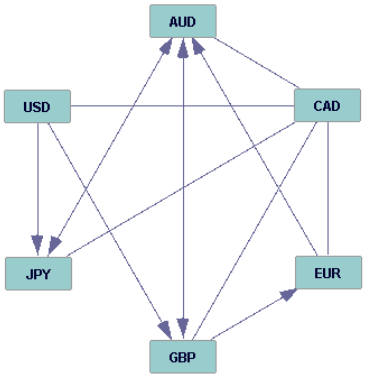
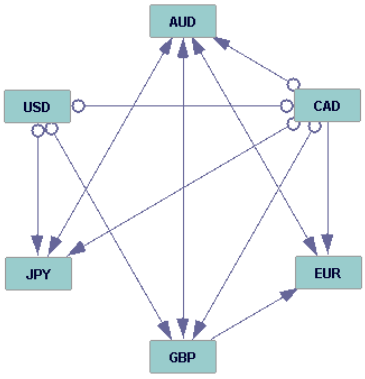


Figure 3.1. FOREX futures market structures, 2009

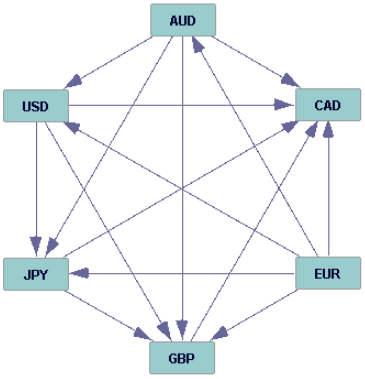
Note: The images show the output of (A) the PC algorithm, (B) the FCI algorithm, and (C) the LiNGAM algorithm applied to the 2009 estimated vector autoregression innovation process $\hat{\boldsymbol{n}}(t)$. These directed graphs represent the contemporaneous causal structure underlying the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the United States dollar (USD) for the year 2009.



A



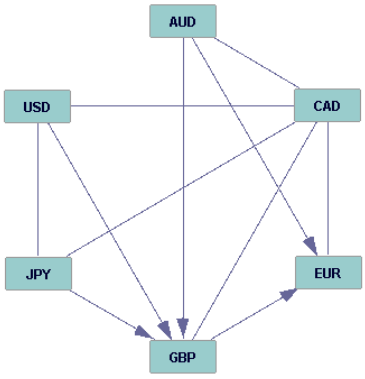
B



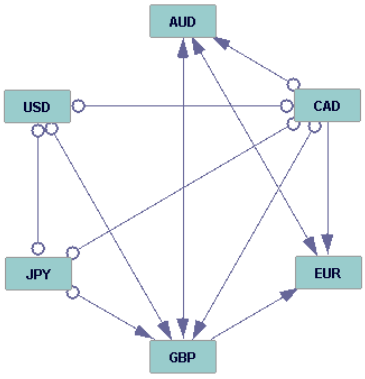
C

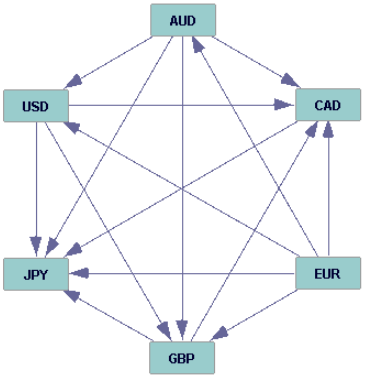
Figure 3.2. FOREX futures market structures, 2010

Note: The images show the output of (A) the PC algorithm, (B) the FCI algorithm, and (C) the LiNGAM algorithm applied to the 2010 estimated vector autoregression innovation process $\hat{\boldsymbol{n}}(t)$. These directed graphs represent the contemporaneous causal structure underlying the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the United States dollar (USD) for the year 2010.



A





B C

Figure 3.3. FOREX futures market structures, 2011
Note: The images show the output of (A) the PC algorithm, (B) the FCI algorithm, and (C) the LiNGAM algorithm applied to the 2011 estimated vector autoregression innovation process $\hat{\boldsymbol{n}}(t)$. These directed graphs represent the contemporaneous causal structure underlying the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the United States dollar (USD) for the year 2011.

2009 Graphical Structure Comparison

The output of the PC algorithm for the year 2009 shows that the graphical structure is thoroughly connected, and that there are five edges that cannot be directed.

Overall, the results of the PC algorithm indicate that there are three possible causal sources (AUD, GBP, and USD) and three possible causal sinks (CAD, EUR, and JPY).

The graphical skeleton resulting from the FCI algorithm is almost the same as that of the PC algorithm, but the interpretation of the edges is quite different. Now AUD cannot be a source because all connected edges are directed into it. Many edges are marked with "o", leaving many undefined ancestor relationships. The FCI algorithm identifies three common causes all connected to AUD.

LiNGAM identifies EUR as the one source and JPY as the one sink. This algorithm identifies edge connection strengths and these are shown in table B-5. The largest causal connection is from USD to JPY measured as 0.75. For the most part, the edges identified by LiNGAM are the same as the other two algorithms. There are some disagreements in terms of edge direction between LiNGAM and the other two algorithms. LiNGAM identifies EUR as a source with edges directed as EUR→AUD and EUR→USD whereas these edges are directed into EUR by the other two algorithms. LiNGAM finds other edges that contradict the direction of those found by the PC algorithm; these are: CAD→AUD, EUR→AUD, EUR→USD, and EUR→GBP.

2010 Graphical Structure Comparison

USD and CAD are the only possible sources in the graphical structures found by the PC and FCI algorithms for the 2010 data. The PC algorithm's output contains two doubly directed edges, which indicates that there is a problem as noted in the algorithm's discussion. This algorithm was also unable to direct any of the edges from CAD.

The FCI algorithm directs four of the five edges away from CAD. It also leaves many of the USD and CAD ancestor relationships undefined. The FCI algorithm again identifies three common causes, all connected to AUD.

LiNGAM once again identifies EUR as the one source, but now CAD is the one sink. The causal connection with the greatest strength in 2010 is EUR→USD, and USD→JPY is also identified as being strong (see table B-5). With the exception of adding AUD→USD and EUR→USD, LiNGAM identified the same edges as the other two algorithms. The disagreements in terms of edge direction with the other two algorithms pertain to LiNGAM identifying EUR as a source node and CAD as a sink.

2011 Graphical Structure Comparison

In 2011, the PC algorithm identifies AUD and CAD as possible sources and CAD and EUR as possible sinks. The FCI algorithm identifies CAD, JPY, and USD as possible sources and AUD and EUR as possible sinks. FCI finds two latent common causes, and both are connected to AUD. LiNGAM identifies EUR as the one source for a third time with JPY as the one sink. Once again, the main difference between the LiNGAM results and the other two algorithms concerns EUR being a source. LiNGAM also finds four more edges than the other algorithms.

Stability Over Time

The results from the PC algorithm are relatively stable over time. The PC algorithm finds three fewer edges in 2010 than it does in 2009, and the 2011graph has

the same number of edges as the 2010 graph but contains the edge JPY→GBP instead of AUD↔JPY. Over time, only the edges GBP−JPY and EUR−AUD completely reverse their orientation, but many of the edges are not oriented one way or another because of the algorithm's uncertainty.

With the exception of the EUR−JPY edge in the 2009 PC algorithm graph, the FCI and PC algorithms find the same graph skeleton in each year. The FCI algorithm identifies only two ancestor relationships according to its edge interpretation rules, and these are CAD→EUR and GBP→EUR in both the 2010 and 2011 graphs. The FCI algorithm identifies four potential common causes, with three at most present in any one year. These potential common causes are all common to the Australian dollar: AUD↔CAD, AUD↔EUR, AUD↔GBP, and AUD↔JPY. The possible shared common cause AUD↔EUR is present in the graphs of all three years. The AUD↔CAD and AUD↔JPY possible common causes are contained in the 2009 and 2010 graphs while AUD↔GBP is shared by the 2010 and 2011 graphs.

In contrast to the PC algorithm, the FCI algorithm produces no conflicting edge orientations in the three sequential graphs, but this is due to the presence of so many unrestricted ("∘") endpoints. The FCI algorithm finds much uncertainty, i.e. many unrestricted endpoints, in the 2009 graph and little of it is resolved in the 2010 or 2011 graphs. Thirteen unrestricted endpoints are found in the 2009 graph, seven in the 2010 graph, and ten in the 2011 graph.

The LiNGAM algorithm finds exactly the same graph skeleton in each year, and EUR remains the one source throughout the time period. The LiNGAM edge

orientations are also consistent through time. The only edges that reverse orientation are the AUD−CAD, AUD−GBP, AUD−USD, and CAD−JPY edges.

Overall, the foreign exchange market structure appears to be quite stable over time. The FCI algorithm found latent common causes, but these do not appear to have any dramatic effect on the foreign exchange market's graphical skeleton structure. Many of the endpoints found by the FCI algorithm were ambiguous in 2009 and 2010, so it is difficult to tell whether or not the latent variables had a significant effect on edge directions. In contrast, most of the edge directions found by LiNGAM are stable over time.

Conclusions

This study uses the PC, FCI, and LiNGAM causal search algorithms to find possible causal structures underlying the foreign currency market. The six most widely traded currencies (Australian dollar, Canadian dollar, euro, Great Britain pound sterling, Japanese yen, and United States dollar) are examined on a 15 minute time frame. Vector autoregressive (VAR) models are applied yearly to the 2009, 2010, and 2011 data of this currency set so that each year is modeled individually. The causal search algorithms use the innovation process from the VAR as input and give a possible graphical model representing the causal structure as output.

The algorithms find that the foreign exchange market's graphical skeleton is relatively stable over time. The most optimistic result, found by the LiNGAM algorithm, is that the foreign exchange market's graphical skeleton is the same in each year.

Edge orientations are also quite consistent through time. In the best case, the FCI algorithm produces no conflicting edge orientations in any of the three years. The FCI algorithm, however, finds much uncertainty, and it reflects this by producing many unrestricted edge endpoints in its graphs. The large number of unrestricted endpoints indicates that there might be many latent variables in the foreign exchange market's graphical structure in addition to the four potential common causes identified by the FCI algorithm; this result reflects the findings of the dark matter literature (Evans 2012).

The possible latent variables and common causes that influence the foreign exchange market potentially make forecasting more difficult than it would be otherwise because there is no way to know how unknown information influences the market. This forecasting difficulty is shown in practice by studies such as Cheung, Chinn, and Pascual (2005). If latent variables are identified and the graphical structure of the foreign exchange market remains stable over time, as its tendency in figures 3.1-3.3, then the inherent stability of the foreign exchange market might make forecasting it easier than forecasting less stable markets.

The FCI algorithm finds many latent variables, and its graph structures are quite different than those of LiNGAM. These findings indicate that it might be potentially dangerous to use VAR-LiNGAM for portfolio management, as described in Chapter II, without first identifying and incorporating the latent variables into the VAR-LiNGAM model. If any of the assumptions of the VAR-LiNGAM model are violated by the presence of the latent variables, such as the random disturbance terms being mutually

independent and temporally uncorrelated, then it would not be advisable to use the model as the basis for portfolio management.

It is clear that much more research is needed to identify the latent variables, if any, which affect the foreign exchange market. Further experience using causal search algorithms to find graphical structures with market data is also needed to determine under what conditions each algorithm produces graphs that adequately describe market structures.

CHAPTER IV

EFFECTS OF THE SWISS FRANC/EURO EXCHANGE RATE FLOOR ON THE CALIBRATION OF PROBABILITY FORECASTS

Introduction

On September 6, 2011 the Swiss National Bank (SNB) began intervening in the Swiss franc/euro (CHF/EUR) exchange rate market to prohibit the franc from appreciating beyond 1.20 francs per euro, and it continued this intervention throughout 2012 (U.S. Department of the Treasury 2011; U.S. Department of the Treasury 2012). The objective of this chapter is to assess the impact of this currency manipulation on the probability forecasts of the CHF/EUR from three time series models. One-step-ahead forecasts of the CHF/EUR probability distribution are based on a series of intraday data for six exchange rates (all versus the Swiss franc). Probability forecasts are generated from a vector autoregression (VAR) model, a VAR model augmented with the LiNGAM causal learning algorithm, and a univariate autoregressive model built on the independent components of an independent component analysis. The forecasted probability distributions are tested for calibration and ranked with two different scoring techniques in periods of time before, surrounding, after, and long after the beginning of the CHF/EUR exchange rate manipulation. A method based on directed acyclic graphs is used as a complement to the scoring metrics in an attempt to identify whether or not the preferred forecasting system changes over time; both the PC and LiNGAM algorithms are used to generate these graphs.

In contrast to other literature on exchange rate forecasting that examines point forecasts of exchange rates, this study follows the example set by Kling and Bessler (1989) and evaluates forecasted probability distributions. A brief summary of the most relevant literature concerning the exchange rate forecasting performance of multivariate time series models is as follows. Liu, Gerlow, and Irwin (1994) determines that the forecasting accuracy of restricted VAR models is better than that of unrestricted VAR models for forecasting the US dollar/yen, US dollar/Canadian dollar, and US dollar/Deutsche mark monthly exchange rates. Hoque and Latif (1993) uses VAR, Bayesian VAR, and vector error correction (VEC) models to forecast the Australian Dollar/United States Dollar monthly exchange rate and concludes that the VEC exhibits superior forecasting performance. Cuaresma and Hlouskova (2005) uses a VAR, restricted VAR, Bayesian VAR, VEC, and Bayesian VEC to forecast five Central and Eastern European monthly exchange rates and conclude that none of the models outperform the others for three month forecasts and that the Bayesian models tend to perform better than the others for five month forecasts. Carriero, Kapetanios, and Marcellino (2009) forecasts the monthly exchange rates of 33 exchange rates against the US dollar using a large Bayesian VAR model; the results indicate that the Bayesian VAR model forecasts better than a random walk model for most of the currencies.

Prequential Analysis

Let $x_t = (x_{1t}, \dots, x_{mt})$ be the observed values of an $m \times 1$ vector time series X_t at time period *t*. Suppose that at any time *n*, the forecaster knows values $x_t,\ t = 1, \dots, n$

and must issue a set of probability distributions P_{n+1} for the next observation X_{n+1}. A prequential forecasting system (PFS) is a rule which associates a choice of P_{n+1} with each value of *n* and with any possible set of outcomes $x_t, t = 1, \ldots, n$ (Dawid 1984). A PFS is so named because it is the combination of probability forecasting and sequential prediction.

Dawid (1984) suggests that the adequacy of a PFS as a probabilistic explanation of the data should depend only on the sequence of forecasts that the PFS in fact made; this is called the prequential principle. In practice, the prequential principle is implemented by using the calibration criterion to judge whether or not a PFS issues adequate probabilities. For a PFS to be well calibrated according to the calibration criterion, the PFS must assign a probability to each event that matches that event's ex post relative frequency.

Formal testing of calibration relies on the probability integral transform as shown in Dawid (1984) and summarized as follows. For a continuous random variable $X_{i,t+1}$ (i.e. the one period forecast for time series *i*), let $U_{i,t+1} = F_{i,t+1}(X_{i,t+1})$ be the continuous distribution function of $P_{i,t+1}$. Under $P_{i,t+1}$ the $U_{i,t+1}$ are independent uniform U[0,1] random variables so that $P_{i,t+1}$ is considered to be well calibrated if the observed sequence of fractiles $u_{i,t+1} = F_{i,t+1}(x_{i,t+1})$ "looks like" a random sample from U[0,1]. In other words, the PFS is well calibrated if the observed sequence $u_{i,t+1} = F_{i,t+1}(x_{i,t+1})$ has cumulative distribution function $G(u_{i,t+1}) = u_{i,t+1}$.

The cumulative distribution function $G(U_{i,t+1})$ for $U_{i,t+1}$ is estimated by arranging the observed sequence $u_{i,t+1} = F_{i,t+1}(x_{i,t+1}),\ t = 1, \ldots, N$ in order of ascending value $u_{i,t+1}(1), \ldots, u_{i,t+1}(N)$ and calculating

$$\hat{G}[u_{i,t+1}(j)] = j/N\,,\ j = 1, \ldots, N. \tag{4.1}$$

Calibration performance can be shown graphically as a plot of the PFS's observed fractiles ($u_{i,t+1}$'s) on the x-axis against the estimated cumulative distribution function $\hat{G}(U_{i,t+1})$ on the y-axis. This calibration plot will be approximately a 45-degree line for a well calibrated PFS.

In practice, a chi-squared goodness-of-fit test can be performed to test a PFS for calibration. This test uses the sequence of observed fractiles ($u_{i,t+1}$'s) from the sequence of probability forecasts $P_{i,t+1}$. Under the null hypothesis that the forecasts are well calibrated, the distribution of a sequence of *N* observed fractiles is a uniform distribution on the interval [0,1], whereas the alternative hypothesis is that the distribution of observed fractiles is not uniform. If the interval [0,1] is divided into *J* nonoverlapping subintervals of length *L* (where $0 \leq L \leq 1$), the goodness-of-fit statistic is calculated as

$$X^2 = \sum_{j=1}^{J} (a_j - L_j N)^2 / L_j N \tag{4.2}$$

where a_j is the actual number of observed fractiles in interval *j* and L_j is the length of interval *j* (Kling and Bessler 1989). The goodness-of-fit statistic is compared to the chi-squared distribution with $J - 1$ degrees of freedom. This test and all other chi-squared goodness-of-fit tests share a common form which is a sum of terms containing the

square of a difference between an observed count and an expected count divided by the expected count

$$\sum (observed - expected)^2 / expected. \quad (4.3)$$

For more information on the goodness-of-fit test see DeGroot and Schervish (2002).

Scoring Forecasts

In addition to calibration plots and calibration tests, prequential forecasting systems can be evaluated by metrics such as the mean-squared error (MSE) criterion or the probability score (Brier 1950). The MSE criterion is most often used to evaluate point forecasts, but it can also be used to evaluate predictive distributions (Kling and Bessler 1989). The MSE is calculated for probability forecasts by using the expected value of the forecast distribution. Let $P_{i,n+1}, n = 1, \ldots, K$ be a sequence of probability forecasts for the *i*th element $X_{i,n+1}$ of the random time series vector X_{n+1} and $M_{i,n+1}$ be the expected value of the distribution $P_{i,n+1}$. The MSE of the forecasts for $X_{i,n+1}$ is calculated as follows

$$MSE = \frac{1}{K}\sum_{n=1}^{K}\left(x_{i,n+1} - M_{i,n+1}\right)^2 \quad (4.4)$$

where $x_{i,n+1}$ is the observed value of $X_{i,n+1}$. The sequence of forecasts with the smallest MSE is preferred; a PFS P is chosen over an alternative PFS Q if the PFS P has the smallest MSE.

In contrast to the MSE, the probability score evaluates the entire forecasted probability distribution (Brier 1950). On any occasion *n+1* suppose that there are *R* possible outcomes for $X_{i,n+1}$ with probabilities $f_{n+1}^{j},\ j = 1, \ldots, R$ so that

$$\sum_{j=1}^{R} f_{n+1}^{j} = 1,\ n = 1, \ldots, K. \tag{4.5}$$

The probability score is defined as

$$PS = \frac{1}{K}\sum_{j=1}^{R}\sum_{n=1}^{K}(f_{n+1}^{j} - E_{n+1}^{j})^2 \tag{4.6}$$

where E_{n+1}^{j} takes the value 1 if outcome j occurred and 0 otherwise. The usage of the probability score is similar to that of the MSE; the sequence of forecasts with the smallest probability score is preferred. A PFS P is chosen over an alternative PFS Q if the PFS P has the smallest probability score.

Directed Acyclic Graphs

A brief overview of graphs is given here because preferred forecasting systems can be identified by using a graphical concept called d-separation and because the LiNGAM algorithm and the PC algorithm (both discussed below) generate output that can be interpreted as a graph representing a set of causal relationships. For a more thorough overview of graphs and their representation of causal relationships, see Pearl (2000).

A directed graph is a diagram that depicts a set of vertices (nodes) V connected by a set of directed edges (links) E with each edge connecting a pair of vertices, e.g. for $\{\mathrm{X}, \mathrm{Y}\} \in \mathrm{V}$, $X \rightarrow Y$. In a directed acyclic graph (DAG), there is no path (sequence of

directed edges) that leads from a vertex back to itself. In other words, a DAG contains no directed cycles, e.g. $X \rightarrow Y \rightarrow X$.

Directed acyclic graphs are often used to represent the causal relationships among a set of random variables. In the causal setting, each vertex represents a variable, and all directed edges are drawn with arrowheads that indicate the direction of causal flow, e.g. $X \rightarrow Y$ indicates that variable X causes variable Y. A set of variables V is causally sufficient if every common cause of two variables in V is also in the set V. A DAG G on a causally sufficient set of variables V is a causal DAG if and only if there is an edge from X to Y in G if and only if X is a direct cause of Y relative to V (Spirtes, Glymour, and Scheines 2001).

Using D-Separation to Identify Preferred Forecasting Systems

A recent paper, Bessler and Wang (2012), suggests that forecasts of preferred economic models d-separate forecasts of less preferred economic models from the actual realizations of the forecasted variable. This notion of model preference is used as a contrast to the MSE and probability score metrics in the application section of this paper.

The notion of d-separation applies to directed acyclic graphs and is fully explained in Pearl (2000). The following definition of d-separation is from Bessler and Wang (2012). A path p defined on a set of forecasts ($f_i,\ i = 1, \ldots, n$) and the actual realization X is said to be d-separated (blocked) by a set of nodes (forecasts) Z if and only if

1) p contains a chain $f_i \rightarrow f_m \rightarrow X$ or a fork $f_i \leftarrow f_m \rightarrow X$, such that the middle node (forecasts from model m) is in Z.
2) p contains an inverted fork (collider) $f_i \rightarrow f_m \leftarrow f_j$ such that the middle node (forecast or actual) m is not in Z and no descendent of m is in Z.

A set Z d-separates f_i from X if and only if Z blocks every path from node f_i (forecast i) to node X (actual). The forecast f_m in the d-separation definition blocks the path between the other forecasts $f_i,\ i \neq m$ and X.

Independent Component Analysis

Independent component analysis is based on the premise that n observed variables $x_1, \dots, x_n$ are linear combinations of underlying, statistically mutually independent source variables $s_1, \dots, s_n$

$$x_i = a_{i1}s_1 + a_{i2}s_2 + \cdots + a_{in}s_n \quad for\ all\ \ i = 1, \dots, n. \tag{4.7}$$

This basic ICA model is written in vector-matrix form as

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s}. \tag{4.8}$$

The observed variables $\boldsymbol{x}$ are used to estimate both the unknown mixing coefficient matrix $\boldsymbol{A}$ and the unobserved independent component vector $\boldsymbol{s}$. The observed variables $\boldsymbol{x}$ and the independent components $\boldsymbol{s}$ are both assumed to have zero mean. If this assumption does not hold then it can be made true by centering the original observed variables, denoted by $\boldsymbol{x}_o$, in a preprocessing step

$$\boldsymbol{x} = \boldsymbol{x}_o - E(\boldsymbol{x}_o) \tag{4.9}$$

This preprocessing also forces the independent components to have zero mean since

$$E(\boldsymbol{s}) = \boldsymbol{A}^{-1}E(\boldsymbol{x}). \quad (4.10)$$

Estimation of the basic ICA model relies on the following assumptions (Hyvarinen, Karhunen, and Oja 2001)

1) The independent components are assumed to be statistically independent, but in application this does not need to be exactly true.
2) The mixing matrix $\boldsymbol{A}$ is assumed to be square and invertible for the sake of convenience and simplicity.
3) The independent components must have non-Gaussian distributions.

Some ICA models are slightly different from the basic ICA model and have their own assumptions; for details see Hyvarinen, Karhunen, and Oja (2001).

The independent components $\boldsymbol{s}$ are more than just uncorrelated; they are as statistically independent as possible. Achieving this requires more information than what is available in a correlation matrix unless all of the variables are normally distributed, in which case zero correlation is equivalent to independence. The estimation of independent components uses higher order moments or other information such as the autocovariance structure for time series variables in addition to correlation information.

It is always possible to linearly transform the observed random variables $\boldsymbol{x}$ into uncorrelated variables. It is also often desirable that the transformed variables have variances equal to unity. The process called whitening transforms zero mean variables $\boldsymbol{x}$ into uncorrelated variables $\boldsymbol{z}$ that have unit variances. The result of whitening is a matrix $\boldsymbol{V}$ that decorrelates the observed data vector

$$\boldsymbol{z} = \boldsymbol{V}\boldsymbol{x}. \tag{4.11}$$

The matrix $\boldsymbol{V}$ is computed as

$$\boldsymbol{V} = \boldsymbol{D}^{-1/2}\boldsymbol{E}^{T} \tag{4.12}$$

where $\boldsymbol{E} = (\boldsymbol{e}_1 \dots \boldsymbol{e}_n)$ is the matrix whose columns are the unit-norm eigenvectors of the covariance matrix $\boldsymbol{C}_{\boldsymbol{x}} = E\{\boldsymbol{x}\boldsymbol{x}^T\}$ and $\boldsymbol{D} = diag(d_1 \dots d_n)$ is the diagonal matrix of the eigenvalues of $\boldsymbol{C}_{\boldsymbol{x}}$. There are an infinite number of matrices $\boldsymbol{V}$ that can create decorrelated components. This is the reason that estimation of the basic ICA model requires the higher order moments of non-Gaussian distributions.

The basic ICA model in equation 4.8 has the following ambiguities

1) The variances of the independent components cannot be determined.
2) The order of the independent components cannot be determined.

The first ambiguity occurs because any scalar multiple of one of the independent components s_i can be cancelled by dividing the corresponding column of the mixing matrix $\boldsymbol{A}$ by that same multiple. The second ambiguity follows from modifying the model with any permutation matrix $\boldsymbol{P}$ and its inverse

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{P}^{-1}\boldsymbol{P}\boldsymbol{s}. \tag{4.13}$$

Now $\boldsymbol{A}\boldsymbol{P}^{-1}$ is the unknown mixing matrix and $\boldsymbol{P}\boldsymbol{s}$ are the independent components in a different order. In applications that use ICA, neither of these ambiguities is important.

ICA Time Series

If the independent components are time series, as opposed to independent random variables in the basic ICA model, then the ICA model takes the following form (Hyvarinen, Karhunen, and Oja 2001)

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t), \quad t = 1, \dots, T \tag{4.14}$$

where t is the time index. Since time series variables have more structure than independent random variables, the time series autocovariances may be used for estimation instead of the higher-order information that is required in the basic ICA model.

The AMUSE algorithm provides one method to estimate the model above (Hyvarinen, Karhunen, and Oja 2001). This algorithm requires the time-lagged covariance matrix in place of the higher-order moments used in the basic ICA model. The time-lagged covariance matrix is computed as

$$\boldsymbol{C}_{\tau}^{\boldsymbol{x}} = E\{\boldsymbol{x}(t)\boldsymbol{x}(t-\tau)^T\} \tag{4.15}$$

where τ is a lag constant, $\tau = 1, 2, 3, \dots$. This matrix contains the autocovariances of each signal and the covariances between signals.

The algorithm is based on the fact that the instantaneous and lagged covariances of $\boldsymbol{s}(t)$ are zero due to independence. Hence, the time-lagged covariance matrix is used to find a matrix $\boldsymbol{B}$ so that all of the instantaneous and lagged covariances of

$$\boldsymbol{y}(t) = \boldsymbol{B}\boldsymbol{x}(t) \tag{4.16}$$

are equal to zero.

The AMUSE algorithm assumes that all of the ICs have autocovariances different from zero and different from each other. This assumption replaces the assumption of the basic ICA model that the independent components must have non-Gaussian distributions.

The AMUSE algorithm uses whitened, zero mean data $\boldsymbol{z}(t)$ as input and generates the separating matrix $\boldsymbol{W}$ as output so that

$$\boldsymbol{W}\boldsymbol{z}(t) = \boldsymbol{s}(t) \tag{4.17}$$

$$\boldsymbol{W}\boldsymbol{z}(t-\tau) = \boldsymbol{s}(t-\tau). \tag{4.18}$$

The time-lagged covariance matrix is modified to be symmetric by the following computation

$$\overline{\boldsymbol{C}}_{\tau}^{\boldsymbol{z}} = \frac{1}{2}[\boldsymbol{C}_{\tau}^{\boldsymbol{z}} + (\boldsymbol{C}_{\tau}^{\boldsymbol{z}})^{T}] \tag{4.19}$$

so that an eigenvalue decomposition on this new symmetric matrix is well defined. The steps of the AMUSE algorithm are as follows (Hyvarinen, Karhunen, and Oja 2001)

1) Center and whiten the observed data $\boldsymbol{x}(t)$ to obtain $\boldsymbol{z}(t)$.
2) Compute the eigenvalue decomposition of the symmetric, time-lagged covariance matrix (equation 4.19) for some time lag τ.
3) The rows of the estimated separating matrix $\widehat{\boldsymbol{W}}$ are given by the eigenvectors.
4) The estimated separating matrix for the unwhitened data $\boldsymbol{x}$ is $\widehat{\boldsymbol{B}} = \widehat{\boldsymbol{W}}\boldsymbol{V}$ where $\boldsymbol{V}$ is defined in equation 4.12.

Time series models are typically built using observed returns, which are represented in vector form by the notation

$$\boldsymbol{R}(t) = \begin{pmatrix} R_1(t) \\ \vdots \\ R_N(t) \end{pmatrix}, \qquad t \in 1, \dots, T \tag{4.20}$$

where $R_i(t)$ is the return on a particular asset $i \in 1, \dots, N$ at time $t \in 1, \dots, T$. In the following discussion, the vector of observed time series variables is the vector of observed returns, i.e. $\boldsymbol{x}(t) = \boldsymbol{R}(t)$. A prequential forecasting system can be created with the independent components by building on the forecasting method described in Popescu (2009). The following procedure is used to create a prequential forecasting system for a set of observed returns

1) Compute the independent components using the estimated separating matrix

$$\boldsymbol{s}(t) = \widehat{\boldsymbol{B}}\boldsymbol{R}(t), \quad t = 1, \dots, T \tag{4.21}$$

2) Model each independent component with an autoregressive (AR) model

$$s_i(t) = c + \sum_{\tau=1}^{k} \varphi_\tau s_i(t-\tau) + \varepsilon_i(t), \quad i \in 1, \dots, N \tag{4.22}$$

where *c* is a constant, *k* is the number of time-delays (lags) of the autoregression, φ_τ are coefficients, and $\varepsilon_i(t)$ is the innovation process.

3) Compute the estimates of the innovation process as follows

$$\varepsilon_i(t) = s_i(t) - c - \sum_{\tau=1}^{k} \varphi_\tau s_i(t-\tau), \quad i \in 1, \dots, N \tag{4.23}$$

and estimate the probability distributions of the innovations with a method such as kernel density estimation. For an overview of kernel density estimation see Bowman and Azzalini (1997).

4) Obtain samples from the estimated probability distributions of the innovations with a sampling technique such as Latin hypercube sampling. A stratified sampling technique such as Latin hypercube sampling is generally more accurate when there are low-probability outcomes, which is likely to be the case in this application (Hardaker et al. 2004).
5) Use the samples of the innovations in conjunction with historical data and parameter estimates to compute the estimated probability distribution for the one-step-ahead independent components using equation 4.22.
6) Finally, transform the samples of the estimated probability distributions of the independent components into estimated probability distributions of the original variables

$$\boldsymbol{x}(t) = \widehat{\boldsymbol{A}}\boldsymbol{s}(t), \quad t = 1, \dots, T. \tag{4.24}$$

LiNGAM Algorithm

The LiGAM algorithm assumes that the observed variables can be arranged in a causal order so that the data generating process can be represented by a DAG, that the value assigned to each variable is a linear function of values assigned to variables positioned earlier in the causal order, that there are no latent common causes, and that the disturbance terms are mutually independent with non-Gaussian distributions and non-zero variances (Shimizu et al. 2006),. The non-Gaussian assumption is important because this allows LiNGAM to estimate the full causal model with no undetermined parameters.

LiNGAM assumes that the observed variables are linear functions of the disturbance variables. When the mean is subtracted from each variable, this is expressed as

$$\boldsymbol{x} = \boldsymbol{B}\boldsymbol{x} + \boldsymbol{e} \tag{4.25}$$

Solving for $\boldsymbol{x}$, this becomes

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{e} \tag{4.26}$$

where $\boldsymbol{A} = (\boldsymbol{I} - \boldsymbol{B})^{-1}$. Equation 4.25 in addition to the assumption that the disturbance terms are independent and have non-Gaussian distributions is the independent component analysis (ICA) model. The ICA model has the following two indeterminancies that must be resolved before a graphical model can be constructed: neither the order nor the scaling of the independent components is defined. LiNGAM resolves both of these issues by permuting and normalizing the ICA output (i.e. the mixing matrix) to obtain a matrix $\boldsymbol{B}$ containing the DAG connection strengths. The graphical representation of this matrix is the causal DAG model.

Because LiNGAM uses the non-Gaussian information contained in the disturbance terms, its output is just one DAG instead of the class of equivalent DAGs found by most causal learning algorithms. As noted earlier, this output includes parameter estimates for the linear model. The LiNGAM procedure is implemented both in MATLAB code provided by Shimizu et al. (2006) and in the TETRAD IV software package provided by Glymour et al. (2012). In the application below, the MATLAB code is used to produce coefficient estimates, and TETRAD IV is used to produce DAG illustrations.

PC Algorithm

The PC algorithm, named after its authors Peter and Clark, is used in this study to recover the graphical structure underlying a set of forecasts and their associated actual realizations (Spirtes, Glymour, and Scheines 2001). The resulting graph is examined to identify the preferred forecasting systems that d-separate the less preferred forecasting systems from the actual realizations of the forecasted variable as discussed in Bessler and Wang (2012).

The PC algorithm assumes that the true causal graph is acyclic and that there are no hidden common causes of any two variables. The Fisher's z test that is often used in implementations of the PC algorithm assumes that each variable has a Normal distribution and that direct causal influences between variables are linear.

The PC algorithm is implemented in two stages (Spirtes, Glymour, and Scheines 2001). In the first stage, the algorithm starts with a complete (fully connected) undirected graph. Then, edges are removed with a sequence of conditional independence tests. For every pair of variables, the algorithm searches for a set of the other variables that renders the pair conditionally independent. If such a set is found then the edge between the pair of variables is removed from the graph. During this edge removal procedure some undirected edges are directed as a precursor to the second stage of the algorithm. Fisher's z is commonly used instead of a test for conditional independence; this tests whether conditional correlations are different from zero. Fisher's z equation is shown below

$$z(\rho(i,j|k),n) = \left[\frac{1}{2}\sqrt{n-|k|-3}\right] ln\left\{\frac{|1+\rho(i,j|k)|}{1-\rho(i,j|k)}\right\} \qquad (4.27)$$

where $\rho(i,j|k)$ is the population correlation between *i* and *j*, *n* is the number of observations, and $|k|$ is the number of variables in set *k*. If *i*, *j*, and *k* are normally distributed and $r(i,j|k)$ is the sample conditional correlation of *i* and *j* given *k* then $z(\rho(i,j|k),n) - z(r(i,j|k),n)$ has a standard normal distribution.

In the second stage of the algorithm, the resulting partially directed graph of the first stage is then directed more fully based on a set of rules. This set of rules is presented in Meek (1995) and Pearl (2000).

The output of the PC algorithm is a graph that may contain three types of edges. A directed edge between a pair of variables indicates that one variable is a direct cause of the other variable. An undirected edge between a pair of variables indicates that the PC algorithm cannot determine which variable causes the other. A doubly directed edge between a pair of variables indicates that one of the algorithm's assumptions has not been met or that some statistical decisions are inconsistent because the sample size is not large enough. The PC algorithm is implemented in the TETRAD IV software package (Glymour et al. 2012).

VAR Models

A vector autoregression (VAR) built using a time series of return observations (equation 4.20) is written as

$$\boldsymbol{R}(t) = \sum_{\tau=1}^{k} \boldsymbol{M}_{\tau}\boldsymbol{R}(t-\tau) + \boldsymbol{n}(t) \tag{4.28}$$

where k is the number of time-delays (lags) of the autoregression, $\boldsymbol{M}_{\tau}$ are $n \times n$ matrices of coefficients, and $\boldsymbol{n}(t)$ is the innovation process.

To find an estimate $\widehat{\boldsymbol{n}}(t)$ of the innovation process, estimate the vector autoregressive model using any least squares method and compute the estimate of the innovation process as

$$\widehat{\boldsymbol{n}}(t) = \boldsymbol{R}(t) - \sum_{\tau=1}^{k} \widehat{\boldsymbol{M}}_{\tau}\boldsymbol{R}(t-\tau). \tag{4.29}$$

In the application below, the VAR model is used as a one-step-ahead prequential forecasting system by using a multivariate normal distribution as the distribution of the innovations $\widehat{\boldsymbol{n}}(t)$. Estimates of the expected value vector and covariance matrix of $\widehat{\boldsymbol{n}}(t)$ are used as parameters of the multivariate normal distribution. The multivariate normal distribution of the innovations is used in equation 4.28 with historical data and parameter estimates to create a probability distribution for the one-step-ahead return vector $\boldsymbol{R}(t)$.

VAR-LiNGAM

LiNGAM can be combined with the VAR model in a specific way so that the VAR model becomes fully identified as described in Hyvarinen et al. (2010); in the following text, this combined model is called VAR-LiNGAM. The VAR-LiNGAM model is a combination of an autoregressive model with time-delays and a structural

equation model, which does not consider the time series structure in data. The autoregressive portion of VAR-LiNGAM is as follows

$$\boldsymbol{R}(t) = \sum_{\tau=1}^{k} \boldsymbol{B}_{\tau}\boldsymbol{R}(t-\tau) + \boldsymbol{e}(t) \tag{4.30}$$

where k is the number of time-delays (lags) of the autoregression, $\boldsymbol{B}_{\tau}$ are $n \times n$ matrices of coefficients, and $\boldsymbol{e}(t)$ is the innovation process. The structural equation portion of VAR-LiNGAM is

$$\boldsymbol{R} = \boldsymbol{B}\boldsymbol{R} + \boldsymbol{e} \tag{4.31}$$

where $\boldsymbol{e}$ is a vector of disturbances and the diagonal of $\boldsymbol{B}$ is defined to be zero.

The complete VAR-LiNGAM model is the combination of equation 4.30 and equation 4.31

$$\boldsymbol{R}(t) = \sum_{\tau=0}^{k} \boldsymbol{B}_{\tau}\boldsymbol{R}(t-\tau) + \boldsymbol{e}(t) \tag{4.32}$$

where k is the number of time-delays (lags) of the autoregression, $\boldsymbol{B}_{\tau}$ are the $n \times n$ matrices containing the causal effects between returns $\boldsymbol{R}(t-\tau)$ with time lag $\tau = 0, \ldots, k$, and $\boldsymbol{e}(t)$ are random disturbances. The $\boldsymbol{B}_{\tau}$ matrices for $\tau > 0$ correspond to effects from the past to the present, while $\boldsymbol{B}_0$ corresponds to instantaneous effects. The VAR-LiNGAM model is based on three assumptions

1) $\boldsymbol{e}(t)$ are mutually independent and temporally uncorrelated, both with each other and over time.
2) $\boldsymbol{e}(t)$ are non-Gaussian.

3) The matrix $\boldsymbol{B}_0$ corresponds to an acyclic graph.

The model is estimated in two stages. First estimate a traditional vector autoregressive model and compute the residuals of the model as described above. Then perform a LiNGAM analysis on the estimate of the innovation process to obtain an estimate of the matrix $\boldsymbol{B}_0$, which is the solution to the instantaneous causal model

$$\hat{\boldsymbol{n}}(t) = \boldsymbol{B}_0\hat{\boldsymbol{n}}(t) + \boldsymbol{e}(t). \tag{4.33}$$

Finally, use $\boldsymbol{B}_0$ to compute $\boldsymbol{B}_\tau$ for $\tau > 0$

$$\widehat{\boldsymbol{B}}_\tau = \left(\boldsymbol{I} - \widehat{\boldsymbol{B}}_0\right)\widehat{\boldsymbol{M}}_\tau \quad \text{for } \tau > 0. \tag{4.34}$$

where $\widehat{\boldsymbol{M}}_\tau$ are estimated coefficient matrices of the VAR model in equation 4.28.

The VAR-LiNGAM model becomes a prequential forecasting system for the one-step-ahead return vector $\boldsymbol{R}(t)$ with the following procedure. Compute an estimate of the independent components $\hat{\boldsymbol{e}}(t)$ from the estimates of the innovations $\hat{\boldsymbol{n}}(t)$

$$\hat{\boldsymbol{e}}(t) = (\boldsymbol{I} - \boldsymbol{B}_0)\hat{\boldsymbol{n}}(t). \tag{4.35}$$

Because there is essentially no stochastic dependence between the independent components, the probability distributions of the individual independent components can be estimated with a univariate estimation method such as kernel density estimation.

Next, obtain samples from the estimated probability distributions of the individual independent components with a sampling technique such as Latin hypercube sampling. Transform the samples of the independent components into samples of the innovations

$$\hat{\boldsymbol{n}}(t) = (\boldsymbol{I} - \boldsymbol{B}_0)^{-1}\hat{\boldsymbol{e}}(t). \tag{4.36}$$

Finally, samples of the innovations in conjunction with historical data and parameter estimates are used to compute the estimated probability distribution for the one-step-ahead return vector $\boldsymbol{R}(t)$ using equation 4.28.

Application

In the remainder of the paper, probability forecasts of the CHF/ EUR exchange rate are generated from the three time series models. Forecast calibration is evaluated with calibration plots and goodness-of-fit calibration tests. The mean-squared error and the probability score metrics are then used to compare the forecasting accuracy of the models. The code used for forecast generation, calibration, and scoring metrics was programmed and executed with MATLAB, Version 7.7 (2012).

Description of the Data

Data is obtained from the Sierra Chart historical intraday futures data service using Sierra Chart software (Sierra Chart 2012). The data set consists of futures contracts that are traded on the CME Group exchange for the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the Swiss franc (CHF). These currencies are chosen because they had the largest market turnover rates in 2010 according to the Triennial Central Bank Survey (Bank for International Settlements 2010).

Sierra Chart software is used to join each currency's future contracts into a single continuous time series for the corresponding currency; for instance, all futures contracts

for the AUD (June 2010, ..., July 2012) were joined in sequence to form a single continuous time series for the AUD. The original data has one minute periodicity and is aggregated across time into fifteen minute intervals so that the resulting data used in this analysis has fifteen minute periodicity.

The time series for the six currencies are converted to be in terms of the Swiss franc (CHF) so that the data used for the analysis consists of observations of the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the United States dollar (USD) all in terms of the Swiss franc (CHF).

Missing data is replaced by the most recent observation in each currency series. Log returns are then computed by taking the natural logarithm and first-differencing the exchange rates (in that order). All log returns in all time periods are stationary based on Dickey Fuller tests.

Recent History of the Swiss Franc

During the second and third quarter of 2011, the SNB became worried that the appreciation of the franc against the euro was hurting the Swiss economy and increasing the risk of deflation. In August, the SNB drove interest rates to nearly zero and flooded the market with liquidity in an attempt to mitigate the fanc's appreciation, but neither of these actions were completely effective. Finally, the franc's appreciation was halted in September when the SNB placed a floor on the CHF/EUR exchange rate. The sequence of SNB actions were as follows (U.S. Department of the Treasury 2011)

- August 3, 2011: the SNB lowered the upper limit of its target range for the three-month Libor to 0-0.25 percent (from 0-0.75 percent).
- August 10, 2011: the SNB announced additional measures to increase liquidity and reduce the appreciation of the franc. These included pumping more liquidity into the Swiss money market and conducting foreign exchange swap transactions (a policy last used in late 2008).
- August 11, 2011: an SNB official said that a temporary peg to the euro was possible
- September 6, 2011: the SNB announced that it was establishing a floor on the CHF/EUR exchange rate (ceiling on the EUR/CHF exchange rate). The franc would not be allowed to appreciate beyond 1.20 francs per euro.

Model Estimation

To analyze forecasts surrounding the establishment of the floor on the CHF/EUR exchange rate, the futures contract time series data is segmented into four two month data sets. These four forecast data sets have corresponding estimation data sets on which estimates of the econometric models are made. The names and descriptions of these four forecast datasets are as follows. In the before data set, the CHF/ EUR exchange rate is unencumbered. The surrounding data set begins on August 11, 2011 when an SNB official announced that a temporary peg was possible; the SNB formally established a floor on the CHF/ EUR exchange rate near the middle of this data set on September 6, 2011. The after data set begins after the floor has been in effect for just more than a

month. The long after data set begins six months after the exchange rate floor has been in place. The exact dates of the forecast data sets and the dates of their accompanying estimation data sets are shown in table 4.1. Expected values of the currency log returns in each forecast data set are shown in table C-1, and covariance matrices of the currency log returns in the estimation and forecast data sets are shown in tables C-2 and C-3. The estimation results for each of the models on all the estimation data sets are reported in tables C-4 through C-7.

Table 4.1. Data Set Starting and Ending Dates

Data Set	Starting Date	Ending Date
Estimation Data Sets		
before	December 11, 2010	June 10, 2011
surrounding	February 11, 2011	August 10, 2011
after	April 11, 2011	October 10, 2011
long after	September 7, 2011	March 6, 2012
Forecast Data Sets		
before	June 11, 2011	August 10, 2011
surrounding	August 11, 2011	October 10, 2011
after	October 11, 2011	December 10, 2011
long after	March 7, 2012	May 6, 2012

Model estimation is performed using SAS, Version 9.2 (2012). The lag lengths for the estimated VAR models are chosen by using the Hannan-Quinn information criterion and the Schwarz's Bayesian criterion (Quinn 1980). For VAR models in all estimation data sets, both the Hannan-Quinn information criterion and the Schwarz's Bayesian criterion are best (most negative) for lag 1. The VAR model estimates of the autoregressive matrices $\boldsymbol{M_1}$ for the estimation data sets are shown in table C-4.

Each VAR-LiNGAM model is built on an estimated VAR model by applying the LiNGAM structural learning algorithm to the VAR model's estimated innovation processes. As evidence that the VAR-LiNGAM non-Gaussian assumption holds on every estimation data set, a Kolmogorov-Smirnov test performed on each currency's corresponding independent factor confirms that the null hypothesis of normality is rejected with p-value less than 0.01 for each factor. The VAR-LiNGAM model estimates of the autoregressive matrices $\boldsymbol{M}_1$ correspond to those of the VAR model and are shown in table C-4. The VAR-LiNGAM model estimates of the causal effect matrices $\boldsymbol{B}_0$ are shown in table C-5.

Independent component analysis is performed on the currency time series and the independent components are modeled with univariate autoregressive processes. The separating matrices $\boldsymbol{B}$ found by the AMUSE algorithm are shown in table C-6. Independent components are computed using the separating matrices as described in equation 4.21. A Kolmogorov-Smirnov test is performed on each independent component to verify the ICA model's non-Gaussian assumption; the test's null hypothesis of normality is rejected with p-value less than 0.01 for each independent component.

The lag lengths for the estimated AR models are chosen by using Schwarz's Bayesian criterion. For the AR models in all estimation data sets, Schwarz's Bayesian criterion is best (most negative) for lag 1. Thus, the independent components are modeled with AR(1) processes whose parameter estimates are shown in table C-7.

Forecast Generation

A multivariate normal distribution is used to model the one-step-ahead probability distribution of the VAR model innovation process. Latin hypercube samples from the multivariate normal distribution in conjunction with the VAR model parameter estimates and historical data are used to compute one-step-ahead probability distributions for the exchange rate returns.

An estimate of the independent factor process of the VAR-LiNGAM model is obtained from its estimated innovation process. Kernel density estimation with a normal probability window is used to estimate the probability distributions of the VAR-LiNGAM independent factor processes. Latin hypercube samples from the independent factor process distributions are transformed into one-step-ahead distributions of the VAR-LiNGAM innovation processes. The innovation process distribution samples plus the VAR-LiNGAM model parameter estimates and historical data are used to compute one-step-ahead probability distributions for the exchange rate returns.

Kernel density estimation with a normal probability window is used to estimate the probability distribution of each AR innovation process. Latin hypercube samples from the innovation process distributions plus the AR model estimates and historical data are used to compute one-step-ahead probability distributions for the independent components. The forecasted probability distributions of the independent components are transformed into forecasted probability distributions of the exchange rate returns as described in equation 4.24.

Sample one-step-ahead cumulative predictive distributions in each of the forecast data sets for the VAR- LiNGAM model are shown in figure 4.1. These sample predictive cdfs are similar to those generated by the VAR and AR models.

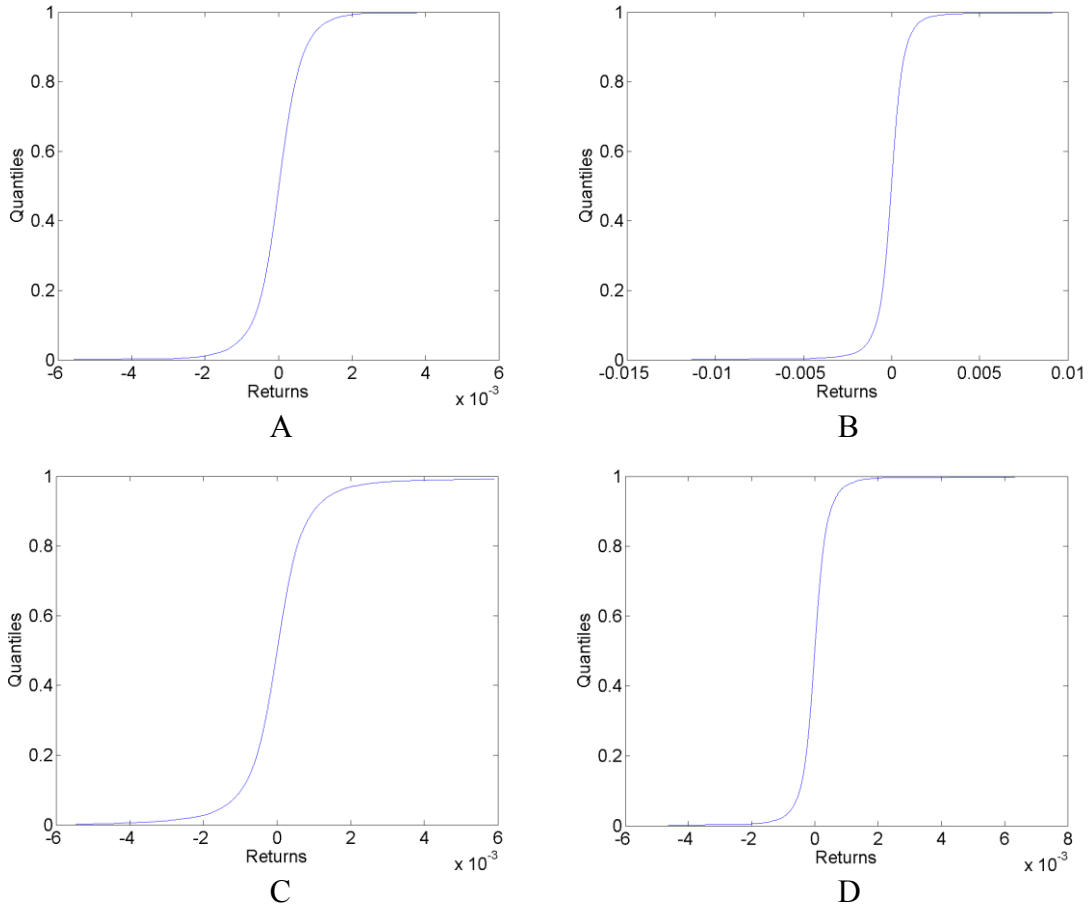
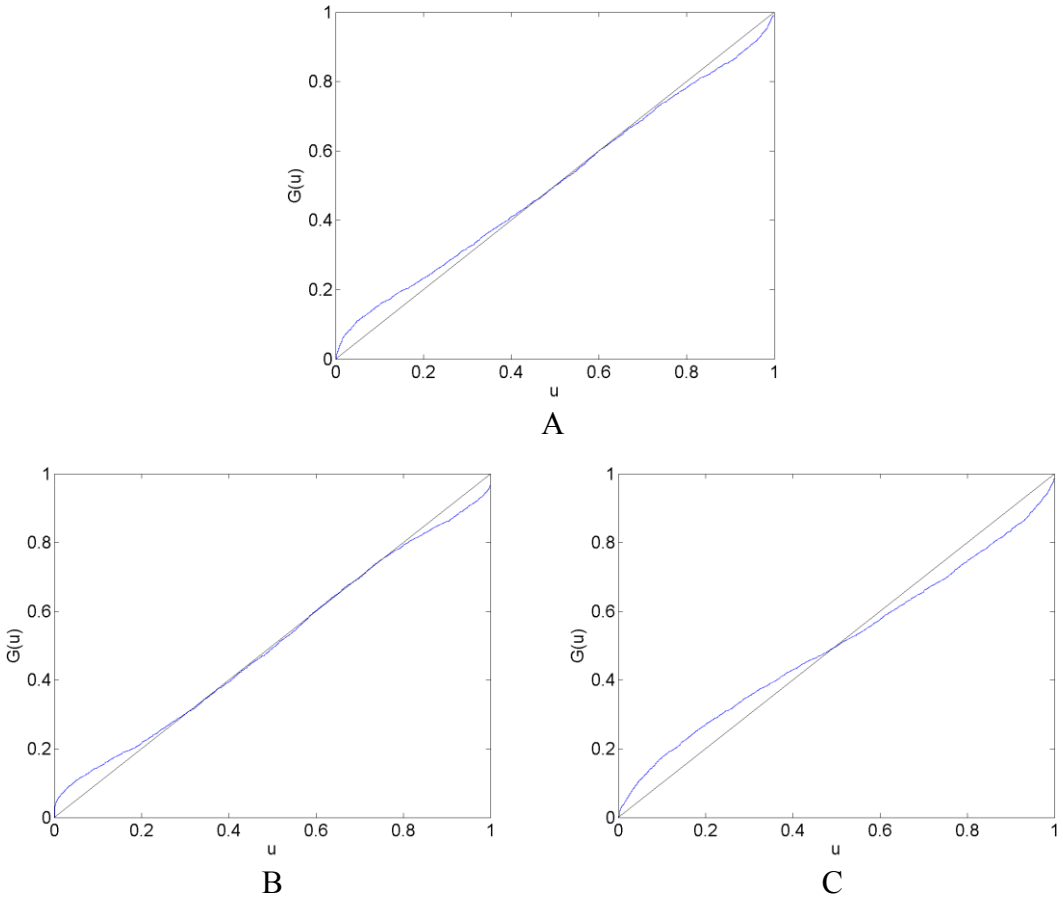


Figure 4.1. Sample cumulative predictive distributions
Note: Sample one-step-ahead cumulative predictive distributions generated by the VAR-LiNGAM model in the before (A), surrounding (B), after (C), and long after (D) forecast data sets.

Forecast Evaluation

The only forecasts considered here are those for the CHF/EUR exchange rate; the forecasts of other currencies are not are evaluated. For the computation of calibration functions, the fractile of each outcome is determined by comparing the outcome to the estimated cumulative predictive distribution. These fractiles are used in conjunction with the estimated cumulative predictive distributions to compute the calibration functions. The calibration functions are both plotted and used to compute goodness-of-fit test statistics.

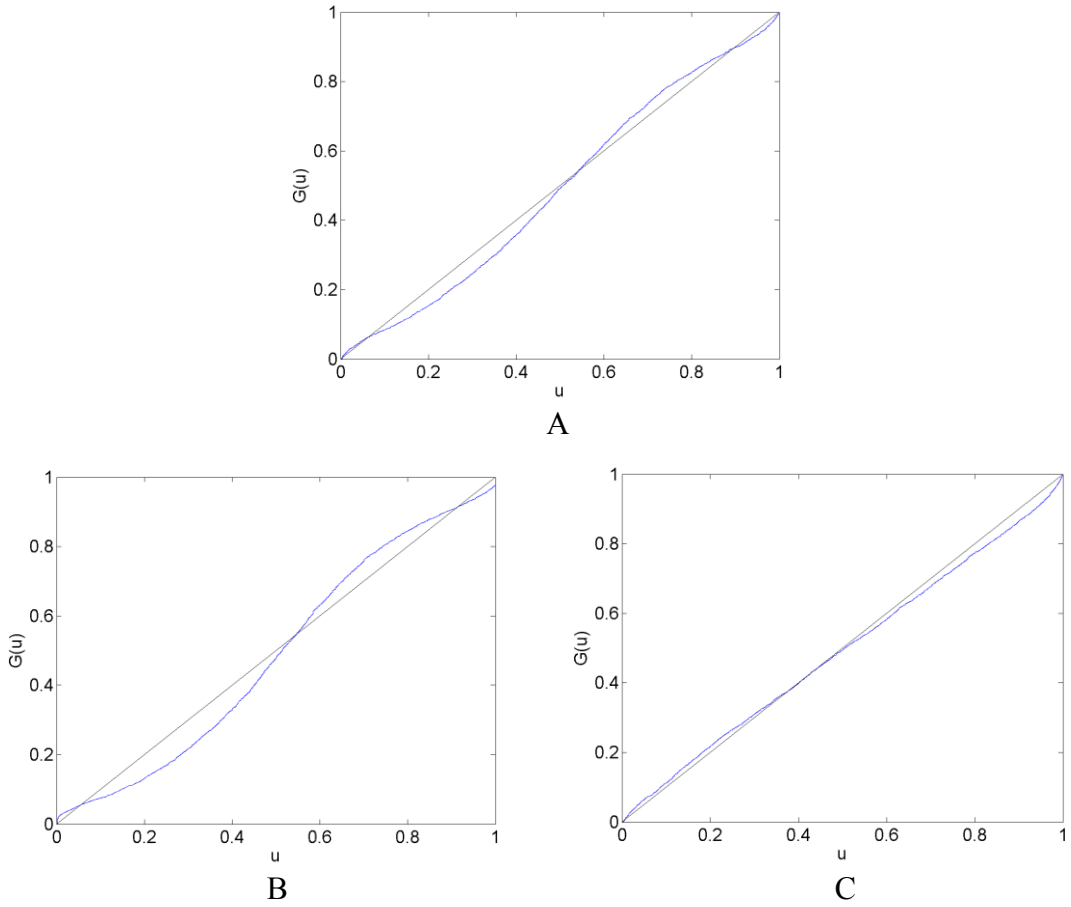
Calibration plots of the CHF/EUR for the before, surrounding, after, and long after forecast data sets are in figures 4.2 through 4.5. The calibration plots for the AR, VAR, and VAR-LiNGAM models in a particular forecast data set in addition to a 45-degree line for reference are shown in each figure. Underconfidence in probability assessments is indicated where the calibration function maps above the 45-degree line, while overconfidence in assessments is indicated where the calibration function maps below the 45-degree line.



A

B C

Figure 4.2. CHF/EUR calibration functions in the before forecast data set
Note: Calibration functions for the CHF/EUR exchange rate are generated by forecasts from the AR (A), VAR (B), and VAR-LiNGAM (C) models in the before forecast data set (June 11, 2011 – August 10, 2011). A model is well calibrated if it maps onto the 45-degree reference line.



A

B

C

Figure 4.3. CHF/EUR calibration functions in the surrounding forecast data set
Note: Calibration functions for the CHF/EUR exchange rate are generated by forecasts from the AR (A), VAR (B), and VAR-LiNGAM (C) models in the surrounding data set (August 11, 2011 – October 10, 2011). A model is well calibrated if it maps onto the 45-degree reference line.

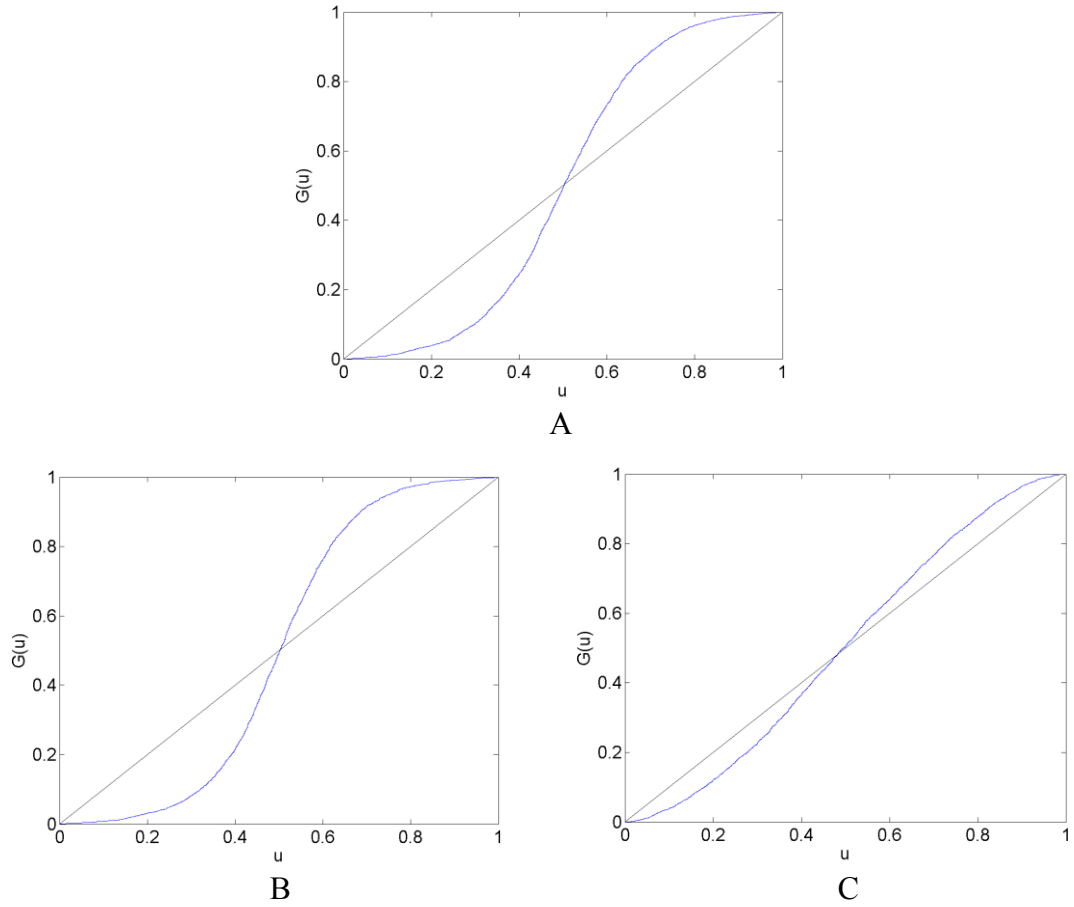


Figure 4.4. CHF/EUR calibration functions in the after forecast data set
Note: Calibration functions for the CHF/EUR exchange rate are generated by forecasts from the AR (A), VAR (B), and VAR-LiNGAM (C) models in the after data set (October 11, 2011 – December 10, 2011). A model is well calibrated if it maps onto the 45-degree reference line.

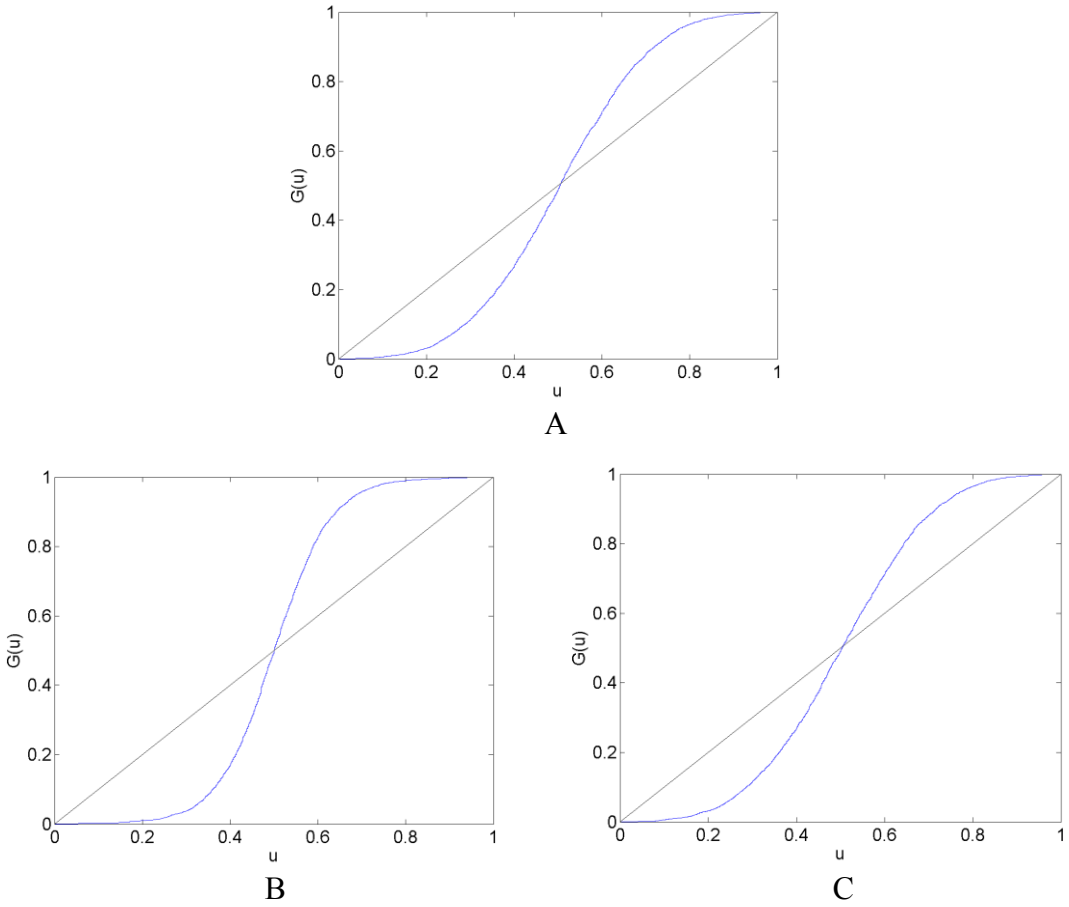


Figure 4.5. CHF/EUR calibration functions in the long after forecast data set
Note: Calibration functions for the CHF/EUR exchange rate are generated by forecasts from the AR (A), VAR (B), and VAR-LiNGAM (C) models in the long after data set (March 7, 2012 – May 6, 2012). A model is well calibrated if it maps onto the 45-degree reference line.

For the before forecast data set, each model exhibits underconfidence on the lower end of the calibration function and overconfidence on the upper end. For the surrounding forecast data set, the AR and VAR models exhibit overconfidence on the lower end of the calibration function and underconfidence on the upper end; the extreme ends of both of these calibration functions show the opposite behavior. The calibration function for the VAR-LiNGAM model on the surrounding data set displays the opposite

behavior of the AR and VAR models with underconfidence on the lower end and overconfidence on the upper end. For the after and long after data sets, the calibration functions for all models exhibit a large degree of overconfidence on the lower end and a large degree of underconfidence on the upper end.

Overall, the calibration plots show that all models are better calibrated (i.e. map closer to the 45-degree line) in the before and surrounding data sets than in the after and long after data sets. Forecasts are less calibrated after the placement of the floor on the CHF/ EUR exchange rate; it appears that the Swiss National Bank's market intervention had a negative affect on the calibration of the time series models in the longer run.

Chi-squared goodness-of-fit tests are performed to test each time series model for calibration during each forecast data set. The null hypothesis that the forecasts are well calibrated is rejected with a p-value near zero in every data set for every time series model; no time series model forecasts are well calibrated in any of the time periods under consideration. Some of the calibration functions appear to map closely to the 45-degree reference line, such as in figures 4.2 A and B. Nevertheless, none of the calibration functions shown in any of figures 4.2 through 4.5 reflect forecasts that are well calibrated according to the goodness-of-fit test.

In some of figures 4.2 through 4.5, the calibration problems appear to be in the tails of the distributions, such as in figures 4.2 A and B. Generating forecasts with distributions estimated via kernel density estimation with a normal probability window might be the source of this bad tail behavior. In the calibration plots that show bad tail behavior, the miscalibration of each tail is in the opposite direction; for example, in

figure 4.2 B, the calibration function shows underconfidence on the low end and overconfidence on the upper end. If the normal probability widow was to blame for this poor tail performance, it would likely produce tails that were too heavy or too light at both ends of the distribution. For instance, if kernel density estimation with a normal probability window produced a distribution with tails that were too light to reflect the distribution of returns then the corresponding calibration function would show underconfidence at both ends of the plot. Additionally, since other figures show that the problem with calibration is more in the central part of the distribution than in the tails, such as figures 4.3 A and B, it is unlikely that the normal probability window is the culprit for bad calibration.

In addition to the calibration tests, the mean-squared error (MSE) and the probability score metrics are used to rank the probability forecasting systems. The mean-squared errors of each model's forecasts are reported in table 4.2, and the probability scores of each model's forecasts are reported in table 4.3. The VAR and VAR-LiNGAM models both have the same MSE on each data set because they are both driven by the innovations of the VAR model (see equation 4.28).

Table 4.2. Mean-Squared Errors

Data Set	AR	VAR &VAR-LiNGAM
before	1.614E-06	1.610E-06
surrounding	2.982E-06	2.983E-06
after	3.409E-07	3.415E-07
long after	2.873E-08	2.722E-08

Note: The table shows the mean-squared errors of the CHF/EUR forecasts from the AR, VAR, and VAR-LiNGAM models on each forecast data set.

Table 4.3. Probability Scores

Data Set	AR	VAR	VAR-LiNGAM
before	0.99876	0.99860	0.99914
surrounding	0.99820	0.99803	0.99877
after	0.99715	0.99713	0.99776
long after	0.99700	0.99697	0.99696

Note: The table shows the probability scores of the CHF/EUR forecasts from the AR, VAR, and VAR-LiNGAM models on each forecast data set.

The MSE results indicate that no model consistently outperforms the others. The VAR and VAR-LiNGAM models perform the best in the before and long after data sets while the AR model performs the best in the surrounding and after data sets. This may indicate that all models have roughly the same forecasting performance or that the VAR and VAR-LiNGAM models perform better in periods isolated from structural change.

In contrast, the probability score rankings show that the VAR model outperforms the other models in all but the long after data set in which the VAR-LiNGAM's performance is slightly better. Because the simple VAR model outperforms the other models that are built using independent components, the probability score results

indicate that there is no gain in forecasting performance when using independent components. Additionally, the probability score ranks the AR forecasts higher than the VAR-LiNGAM forecasts in all periods but the last; this may indicate that in some cases the multivariate VAR-LiNGAM model provides no advantage over the univariate AR model.

Finally, d-separation is used in an attempt to identify the preferred forecasting system as described in Bessler and Wang (2012). A graphical structure is found for each dataset's forecasts and actual realizations using both the PC algorithm and the LiNGAM algorithm. The Tetrad "Knowledge" component is used for both algorithms to prohibit the actual realizations from causing the forecasts (i.e. the future cannot cause the past). The expected values of the forecast distributions and the corresponding actual realizations are used as inputs to both graphical structure search algorithms.

The graphical structures found by the search algorithms are shown in figures 4.6 through 4.9; the VAR and VAR-LiNGAM models are both represented by the node labeled "VAR" in these figures because both models produce identical forecasts of the mean exchange rate. The only d-separation that occurs is in the long after dataset's graphical structure recovered by the PC algorithm (figure 4.9) in which VAR and VAR-LiNGAM forecasts d-separate AR forecasts from the actual realizations. Thus, the VAR and VAR-LiNGAM models are preferred to the AR model in the long after dataset according to the PC algorithm. This result agrees with the rankings of these models by the MSE criterion on the long after dataset. There is no d-separation in any of the other graphical structures as shown in figures 4.6 through 4.8.

The VAR and VAR-LiNGAM models generate better forecasts in the long after period according to the MSE, probability score, and d-separation. This is some indication that the VAR-LiNGAM model performs better than the AR model after market intervention has been in effect for some period of time.

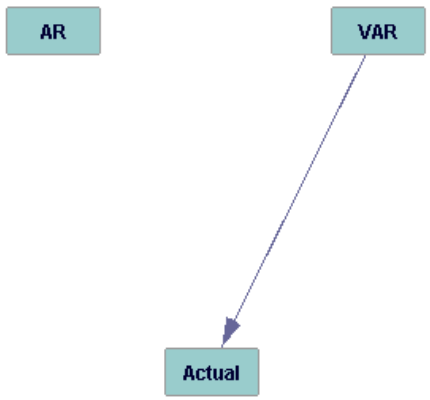


Figure 4.6. DAG from the PC algorithm on forecasts of the CHF/EUR exchange rate and actual outcomes in the before data set
Note: This is the pattern from the PC algorithm on forecasts of the CHF/EUR exchange rate and its Actual outcomes in the before data set (June 11, 2011 - August 10, 2011). Forecasts are of the mean CHF/EUR exchange rate from the AR, VAR, and VAR-LiNGAM models. The VAR and VAR-LiNGAM models are both represented by the node labeled "VAR" because both models produce identical forecasts of the mean exchange rate.

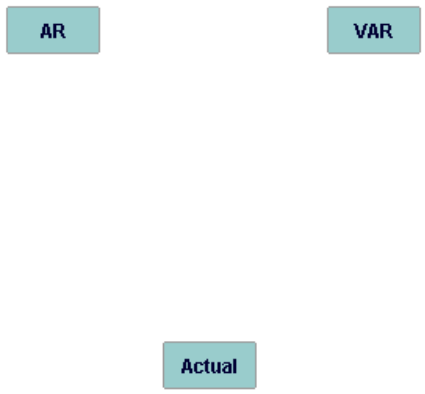


Figure 4.7. DAG showing the independence of forecasts of the CHF/EUR exchange rate and actual outcomes in various data sets

Note: Forecasts of the CHF/EUR exchange rate from all models are independent of the Actual outcomes in the before data set (June 11, 2011 - August 10, 2011) according to the LiNGAM algorithm. Forecasts and Actuals are also independent in the surrounding data set (August 11, 2011 - October 10, 2011) and in the after data set (October 11, 2011 - December 10, 2011) according to both the PC and LiNGAM algorithms. Forecasts are of the mean CHF/EUR exchange rate from the AR, VAR, and VAR-LiNGAM models. The VAR and VAR-LiNGAM models are both represented by the node labeled "VAR" because both models produce identical forecasts of the mean exchange rate.

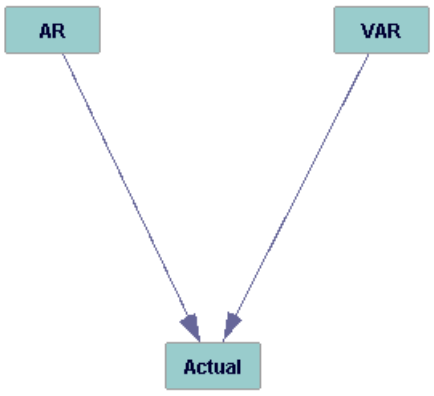


Figure 4.8. DAG from the LiNGAM algorithm on forecasts of the CHF/EUR exchange rate and actual outcomes in the long after data set

Note: This is the directed acyclic graph from the LiNGAM algorithm on forecasts of the CHF/EUR exchange rate and its Actual outcomes in the long after data set (March 7, 2012 - May 6, 2012). Forecasts are of the mean CHF/EUR exchange rate from the AR, VAR, and VAR-LiNGAM models. The VAR and VAR-LiNGAM models are both represented by the node labeled "VAR" because both models produce identical forecasts of the mean exchange rate.

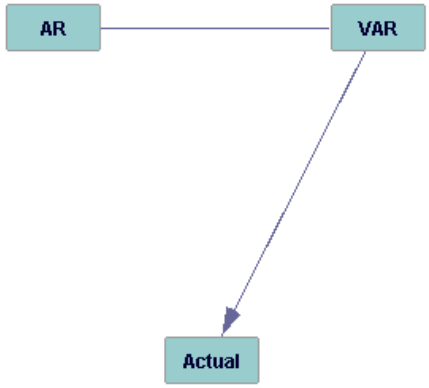


Figure 4.9. DAG from the PC algorithm on forecasts of the CHF/EUR exchange rate and actual outcomes in the long after data set
Note: This is the pattern from the PC algorithm on forecasts of the CHF/EUR exchange rate and its Actual outcomes in the long after data set (March 7, 2012 - May 6, 2012). The graph shows that forecasts from the VAR and VAR-LiNGAM models d-separate forecasts from the AR model and actual outcomes. Forecasts are of the mean CHF/EUR exchange rate from the AR, VAR, and VAR-LiNGAM models. The VAR and VAR-LiNGAM models are both represented by the node labeled "VAR" because both models produce identical forecasts of the mean exchange rate.

Conclusions

This study assesses the impact of the Swiss National Bank's manipulation of the CHF/EUR exchange rate on the probability forecasts from a VAR model, a VAR model augmented with the LiNGAM causal learning algorithm, and a univariate AR model built on the independent components of an independent component analysis. Forecasts are divided among data sets that represent periods of time before, surrounding, after, and long after the beginning of the CHF/EUR exchange rate manipulation.

Calibration plots are shown for the forecasted probability distributions of CHF/EUR returns on all data sets. None of the forecasted probability distributions

appear to be calibrated based on the calibration plots, and calibration tests confirm this. The calibration plots show that all models are better calibrated in the periods before and surrounding the beginning of the exchange rate manipulation than in the two periods after the floor on the CHF/EUR was established. This implies that the SNB‘s intervention in the CHF/EUR market had a negative impact on the forecasting performance of the time series models.

The mean-squared error (MSE) and the probability score metrics are used to rank the probability forecasting systems. When comparing models within each data set, the MSE finds that the VAR and VAR-LiNGAM models generate better forecasts in the before and long after data sets while the AR model generates better forecasts in the surrounding and after data sets. These results may indicate that all models have roughly the same forecasting performance or that the VAR and VAR-LiNGAM models perform better in periods isolated from structural change.

The probability score finds that the VAR model outperforms the other models in all data sets except the long after dataset in which the VAR-LiNGAM’s performance is slightly better. The relatively good performance of the VAR model, which does not take independent components into account, may indicate that there is no improvement in forecasting performance when independent components are used to generate forecasts. Additionally, the probability score ranks the AR forecasts higher than the VAR-LiNGAM forecasts in all periods but the last; this may indicate that in many cases the univariate independent component AR model provides as good or better forecasts than the multivariate VAR-LiNGAM model.

Graphical structures underlying each dataset's forecasts and actual realizations are found using the PC and LiNGAM algorithms. D-separation is found only in the graphical structure recovered by the PC algorithm for the long after dataset. In this structure, the VAR and VAR-LiNGAM forecasts d-separate AR forecasts from the actual realizations indicating that the VAR and VAR-LiNGAM models are preferred to the AR model. This is some indication that the VAR-LiNGAM model performs better than the AR model after market intervention has been in effect for some period of time.

The findings of this study raise some interesting questions. In particular, does central bank intervention in a currency market always have a negative impact on the forecasting performance of time series models? Does the VAR model often generate forecasts that are as good as those from models that use independent components? Under what circumstances does the univariate independent component AR model generate better forecasts than the multivariate VAR-LiNGAM model? These are questions for future studies.

CHAPTER V

SUMMARY

This dissertation applies the combination of time series models and causal search algorithms to the foreign exchange markets to find causal linkages between the six most widely traded currencies (Australian dollar, Canadian dollar, euro, Great Britain pound sterling, Japanese yen, and United States dollar). This information is used to visualize causal connections, enhance the forecasting ability of time series models, and improve portfolio management.

A method for decomposing portfolio risk based on a portfolio's underlying independent risk factors is presented in Chapter II. The decomposition of a risk measure as a summation of the risk contributions from independent factors is introduced and is named total risk decomposition. The independent factor risk contributions are named total risk contributions because they sum to the total portfolio's risk.

Total risk contributions are simpler to compute and easier to interpret than marginal risk contributions, which are traditionally used in portfolio analysis. Each total risk contribution is non-negative and can be interpreted as having a potentially adverse effect on the portfolio. Furthermore, in the total risk contribution framework, each position contributes independently to the risk of the total portfolio so that changing the size of one position does not affect the risk contributions of the other positions. This allows a portfolio manager to know exactly how much risk each position (in an independent factor) will contribute to a portfolio.

When a set of securities conforms closely to the assumptions of independent component analysis, the independent factors can be thought of as the true sources of risk and return. In this case, the total risk contribution framework is recommended for use in both portfolio construction (e.g. optimization) and portfolio analysis because of its simplicity and conceptual clarity.

In an application of this technique to the six most widely traded currencies, a portfolio's underlying independent risk factors are found empirically using the LiNGAM causal search algorithm combined with a VAR time-series model. The mean-variance model is used for portfolio optimization, and its risk metric, the portfolio variance, is decomposed into its independent risk contributions.

Six portfolios are constructed using the canonical mean-variance model with added constraints. The following are constrained during the mean-variance optimization in an attempt to increase portfolio diversification: relative portfolio weights, independent factor relative portfolio weights, percentage marginal risk contributions, independent factor percentage marginal risk contributions, and independent factor total risk contributions.

An efficient frontier is generated for each of the portfolios and a comparison is made of the loss of efficiency due to each of the constraint sets. The relative portfolio weight constraint set produces an efficient frontier that is a large distance from the other efficient frontiers; this indicates that the cost of the relative portfolio weight constraint set is much greater than the cost of the other constraint sets.

A comparison of the internal characteristics of the portfolios on a small out-of-sample data set shows that in addition to the independent factor total risk contribution constraint set, the independent factor relative weight constraint and marginal risk contribution of independent factor constraint sets do a moderate job of controlling total risk contributions. This analysis also indicates that weight constraints, which are a popular way to control portfolio exposures, can be very ineffective at controlling total risk contributions. The overall indication is that the only way to effectively control independent risk contributions is to put constraints on the independent risk contributions themselves.

This study uses a very small data set and focuses exclusively on the foreign exchange market. The methods for controlling independent risk contributions must be deployed on more data and on many different markets before the results of this study can be generalized. If other applications produce favorable results then perhaps controlling independent risk contributions will be useful for constructing actual portfolios.

Three different causal search algorithms (PC, FCI, and LiNGAM) in conjunction with vector autoregression time series models are used in Chapter III to find possible causal structures underlying the foreign currency market in the years 2009, 2010, and 2011. This modeling process results in a directed acyclic graph that illustrates the contemporaneous causal linkages between currencies. The three causal search algorithms do not necessarily find the same causal structure when given the same data set as input.

The causal structure underlying the six most widely traded currencies on a 15 minute time frame is found for each year (2009, 2010, and 2011) by each algorithm. All three of the algorithms find that the foreign exchange market's graphical skeleton is relatively stable over time. While all of the algorithms find stability, LiNGAM finds that the foreign exchange market's graphical skeleton is the same in each year. In addition, all three algorithms find that edge orientations are also quite consistent through time.

The FCI algorithm produces many unrestricted edge endpoints in its graphs. This large number of unrestricted endpoints indicates that there might be many latent variables that affect the six most widely traded currencies in addition to four potential common causes identified by the FCI algorithm.

The presence of latent variables makes it potentially dangerous to use VAR-LiNGAM for portfolio management, as described in Chapter II, without first identifying and incorporating the latent variables into the VAR-LiNGAM model. If any of the assumptions of the VAR-LiNGAM model are violated by the presence of the latent variables, such as the random disturbance terms being mutually independent and temporally uncorrelated, then it would not be advisable to use the model as the basis for portfolio management.

In future studies of the foreign exchange market's causal structure, a larger data set will be required to more adequately determine the market's graphical skeleton stability characteristics (e.g. the longest and shortest durations of stability). The use of higher and lower frequency data sets might also give other insights into the stability of the foreign exchange market at different periodicities.

More experience with the causal search algorithms is needed to determine under what conditions each algorithm gives the most accurate results. Additionally, more diverse data is needed to identify the latent variables identified by the FCI algorithm and their affects on the foreign exchange market.

The impact of the Swiss National Bank's floor on the Swiss franc/euro (CHF/EUR) exchange rate in 2011-2012 on the probability forecasts of the CHF/EUR from three time series models is assessed in Chapter IV. One-step-ahead forecasts of the CHF/EUR probability distribution are based on a series of intraday futures data for the six most widely traded currencies all in terms of the Swiss franc. Probability forecasts of the CHF/EUR are generated from a vector autoregression model, a VAR model augmented with the LiNGAM causal learning algorithm, and a univariate autoregressive model built on the independent components of an independent component analysis. Forecasts are divided among data sets that represent periods of time before, surrounding, after, and long after the beginning of the floor on the CHF/EUR exchange rate.

None of the forecasted probability distributions of the CHF/EUR returns appear to be calibrated on any data set based on calibration plots, and calibration tests confirm this. The calibration plots show that all models are better calibrated in the periods before and surrounding the implementation of the exchange rate floor than in the two following periods in which the floor is in effect. This implies that the floor on the CHF/EUR exchange rate had a negative impact on the forecasting performance of all of the time series models.

The mean-squared error (MSE) and the probability score metrics are used to rank the probability forecasting systems. The MSE finds that the VAR and VAR-LiNGAM models generate better forecasts in the before and long after data sets while the AR model generates better forecasts in the surrounding and after data sets. These results may indicate that all models have roughly the same forecasting performance or that the VAR and VAR-LiNGAM models perform better in periods isolated from structural change.

The probability score finds that the VAR model outperforms the other models in all data sets except the last in which the VAR-LiNGAM's performance is slightly better. The relatively good performance of the VAR model, which does not take independent components into account, may indicate that there is no improvement in forecasting performance when independent components are used to generate forecasts. Additionally, the probability score ranks the AR forecasts higher than the VAR-LiNGAM forecasts in all periods but the last; this may indicate that in many cases the univariate independent component AR model provides as good or better forecasts than the multivariate VAR-LiNGAM model.

To complement the results of the score metrics, a graphical notion called d-separation is used in an attempt to identify the preferred forecasting system. Graphical structures underlying each dataset's forecasts and actual realizations are found using the PC and LiNGAM algorithms. D-separation is found only in the graphical structure recovered by the PC algorithm for the long after dataset. In this structure, the VAR and VAR-LiNGAM forecasts d-separate AR forecasts from the actual realizations indicating

that the VAR and VAR-LiNGAM models are the preferred to the AR model. This result agrees with the rankings of the models in the long after dataset by the MSE criterion.

This study found evidence that intervention in a market has a negative impact on the forecasting performance of time series models. To determine whether this is an isolated result or a general effect of intervention, time-series forecasting performance surrounding other interventions should be considered in not only the foreign exchange market but other markets as well. Other studies might also attempt to determine whether the effects of a short term shock instead of a long term intervention cause a temporary or sustained deterioration of time-series forecasting performance.

There is no one time series model that generates the best forecasts according to all performance metrics in all data sets. Specifically, VAR model forecasts are found to be very competitive to those from models that use independent components, and the univariate AR model built on independent components sometimes generates better forecasts than the multivariate VAR-LiNGAM model. More research is needed to determine the circumstances under which each time series model generates the best forecasts.

Finally, future studies could follow the work of Kling and Bessler (1989) and recalibrate badly calibrated probability forecasts in an attempt to improve the probability forecasting performance in future time periods.

REFERENCES

Akleman, D. G., Bessler, D. A., and Burton, D. M. 1999. "Modeling Corn Exports and Exchange Rates with Directed Graphs and Statistical Loss Functions." In Glymour, C., and Cooper, G. F., eds. *Computation, Causation, and Discovery*. Menlo Park, CA: MIT Press, pp. 497–520.

AMPL. 2011. Albuquerque, NM: AMPL Optimization LLC.

Bank for International Settlements. 2010. *Triennial Central Bank Survey Report on Global Foreign Exchange Market Activity in 2010*. Basel, Switzerland: Bank for International Settlements.

Bessler, D. A., and Wang, Z. 2012. "D-Separation, Forecasting, and Economic Science: A Conjecture." *Theory and Decision*. 73(2):295-314.

Bowman, A. W., and Azzalini, A. 1997. *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*. New York: Oxford University Press.

Brier, G. W. 1950, January. "Verification of Forecasts Expressed in Terms of Probability." *Monthly Weather Review*. 78(1):1-3.

Carriero, A., Kapetanios, G., and Marcellino, M. 2009. "Forecasting Exchange Rates with a Large Bayesian VAR." *International Journal of Forecasting*. 25(2):400-417.

Cheung, Y. W., Chinn, M. D., and Pascual, A. G. 2005. "Empirical Exchange Rate Models of the Nineties: Are Any Fit to Survive?" *Journal of International Money and Finance*. 24(7):1150-1175.

Cuaresma, J. C., and Hlouskova, J. 2005. "Beating the Random Walk in Central and Eastern Europe." *Journal of Forecasting.* 24(3):189-201.

Dawid, A. P. 1984. "Present Position and Potential Developments: Some Personal Views: Statistical Theory: The Prequential Approach." *Journal of the Royal Statistical Society*. Series A:278-292.

DeGroot, M. H., and Schervish, M. J. 2002. *Probability and Statistics*. Boston, MA: Addison-Wesley.

Evans, M. D. 2012. “Exchange-Rate Dark Matter.” Working paper, Department of Economics, Georgetown University.

Frankel, J. A., and Rose, A. K. 1994. “A Survey of Empirical Research on Nominal Exchange Rates.” Working paper. Center for International and Development Economics Research, University of California at Berkeley.

Gerard, C., and Tutuncu, R. 2007. *Optimization Methods in Finance*. Cambridge: Cambridge University Press.

Glymour, C., Scheines, R., Spirtes, P., and Ramsey, J. 2012. The TETRAD Project. Retrieved from http://www.phil.cmu.edu/projects/tetrad/

Gnuplot. 2012. Retrieved from: http://www.gnuplot.info/

Graphviz. 2012. Retrieved from: http://www.graphviz.org/

Green, R. C. 1992. “When Will Mean-Variance Efficient Portfolios be Well Diversified?” *Journal of Finance*. 47(5):1785-1809.

Haigh, M. S., and Bessler, D. A. 2004. “Causality and Price Discovery: An Application of Directed Acyclic Graphs.” *Journal of Business*. 77(4):1099–1121.

Hardaker, J. B., Huirne, R. B., Anderson, J. R., and Lien, G. 2004. *Coping with Risk in Agriculture*. Cambridge, MA: CABI Publishing.

Hoque, A., and Latif, A. 1993. “Forecasting Exchange Rate for the Australian Dollar Vis-a-Vis the US Dollar Using Multivariate Time-Series Models.” *Applied Economics*. 23(5):403-407.

Hyvarinen, A. S. 2008. “Causal Modelling Combining Instantaneous and Lagged Effects: an Identifiable Model Based on Non-Gaussianity.” In *Proceedings of the 25th International Conference on Machine Learning*. Helsinki, Finland: The International Machine Learning Society, pp. 424-431.

Hyvarinen, A., Karhunen, J., and Oja, E. 2001. *Independent Component Analysis*. New York, NY: John Wiley and Sons, Inc.

Hyvarinen, A., Zhang, K., Shimizu, S., and Hoyer, P. O. 2010. “Estimation of a Structural Vector Autoregression Model Using Non-Gaussianity.” *Journal of Machine Learning Research*. 11:1709-1731.

Kim, J. W., Leatham, D. J., and Bessler, D. A. 2007. "REITs' Dynamics Under Structural Change with Unknown Break Points." *Journal of Housing Economics*. 16(1):37–58.

Kling, J. L., and Bessler, D. A. 1989. "Calibration-Based Predictive Distributions: An Application of Prequential Analysis to Interest Rates, Money, Prices, and Output." *The Journal of Business*. 64(4):477-499.

Kuhl, M. 2010. "Bivariate Cointegration of Major Exchange Rates, Cross-Market Efficiency and the Introduction of the Euro." *Journal of Economics and Business*. 62(1):1-19.

Litterman, R. 1996. "Hot Spots and Hedges." *Risk Management Series*, New York, NY: Goldman, Sachs & Co.

Liu, T.-R., Gerlow, M. E., and Irwin, S. H. 1994. "The Performance of Alternative VAR Models in Forecasting Exchange Rates." *Journal of Forecasting*. 10(3):419-433.

Markowitz, H. 1952. "Portfolio Selection." *Journal of Finance*. 7(1):77-91.

MATLAB, Version 7.7. 2012. Natick, MA: The Mathworks Inc.

Meek, C. 1995. "Causal Inference and Causal Explanation with Background Knowledge." In Besnard, P., and Hanks, S., eds. *Proceedings of Eleventh Conference on Uncertainty in Artificial Intelligence*, San Francisco, CA: Morgan Kaufmann, pp. 403-410.

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

Meucci, A. 2007. "Risk Contributions from Generic User-Defined Factors." *The Risk Magazine*, June, pp. 84-88.

Moneta, A. 2008. "Graphical Causal Models and VARs: An Empirical Assessment of the Real Business Cycles Hypothesis." *Empirical Economics*. 35(2):275-300.

Norrbin, S. C. 1996. "Bivariate Cointegration Among European Monetary System Exchange Rates." *Applied Economics*. 28(12):1505-1513.

Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

Phengpis, C. 2006. "Market Efficiency and Cointegration of Spot Exchange Rates During Periods of Economic Turmoil: Another Look at European and Asian Currency Crises." *Journal of Economics and Business*. 58(4):323–342.

Popescu, T. D. 2009. "Time Series Forecasting Using Independent Component Analysis." *World Academy of Science, Engineering and Technology*. 25:667-672.

Qian, E. 2006. "On the Financial Interpretation of Risk Contribution: Risk Budgets Do Add Up." *Journal of Investment Management*. 4(4):1-11.

Quinn, B. G. 1980. "Order Determination for a Multivariate Autoregression." *Journal of the Royal Statistical Society*. 42(2): 182-185.

Rapp, T. A., and Sharma, S. C. 1999. "Exchange Rate Market Efficiency: Across and within Countries." *Journal of Economics and Business*. 51(5):423–439.

SAS, Version 9.2. 2012. Cary, NC: SAS Institute Inc.

Shaffer, J. 1995. "Multiple Hypothesis Testing." *Annual Review of Psychology*. 46(1):561–584.

Shimizu, S., Hoyer, P. O., Hyvarinen, A., and Kerminen, A. 2006. "A Linear Non-Gaussian Acyclic Model for Causal Discovery." *Journal of Machine Learning Research*, 7:2003-2030.

Sierra Chart. 2012. Retrieved from: http://www.sierrachart.com/

Spirtes, P., Glymour, C., and Scheines, R. 2001. *Causation, Prediction, and Search*. Cambridge, MA: MIT Press.

U.S. Department of the Treasury. 2011. *Report to Congress on International Economic and Exchange Rate Policies*. Washington DC, December 27.

U.S. Department of the Treasury. 2012. *Report to Congress on International Economic and Exchange Rate Policies*. Washington DC, November 27.

Zenios, S. A. 2007. *Practical Financial Optimization: Decision Making for Financial Engineers*. Cambridge, UK: Blackwell.

APPENDIX A

CHAPTER II FIGURES AND TABLES

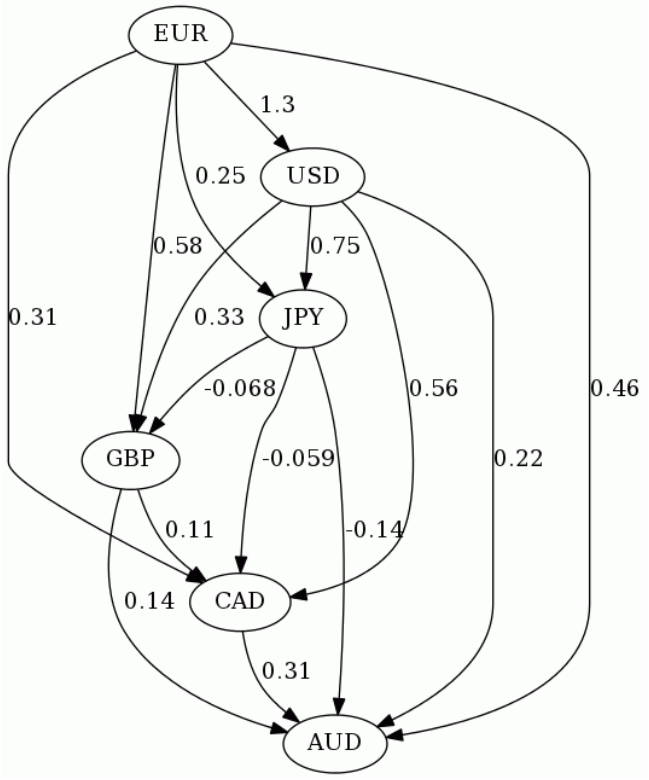


Figure A-1. FOREX market structure, 2009
Note: This DAG shows the instantaneous (lag zero) causal market structure in 2009 of the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the United States dollar (USD). The numeric labels on the graph's edges correspond to coefficients in the 2009 instantaneous causal effect matrix $\mathbf{B}_0$.

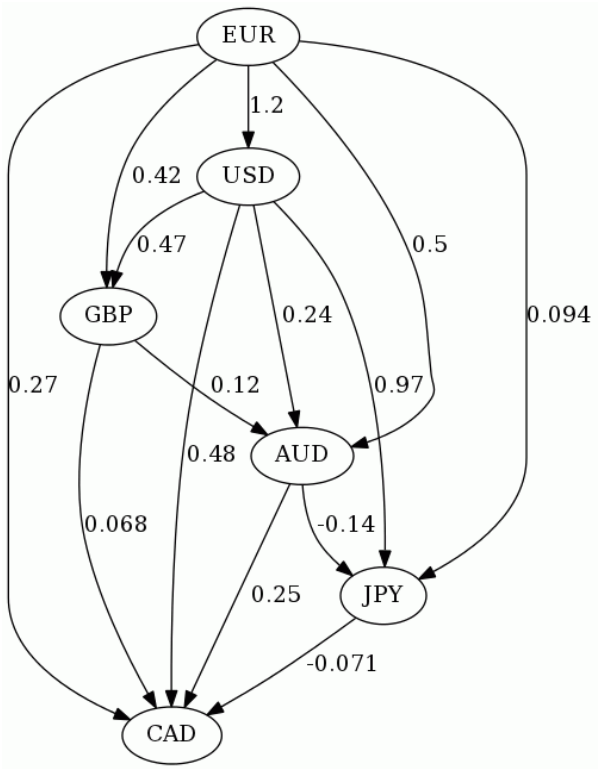


Figure A-2. FOREX market structure, 2010

Note: This DAG shows the instantaneous (lag zero) causal market structure in 2010 of the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the United States dollar (USD). The numeric labels on the graph's edges correspond to coefficients in the 2010 instantaneous causal effect matrix $\mathbf{B}_0$.

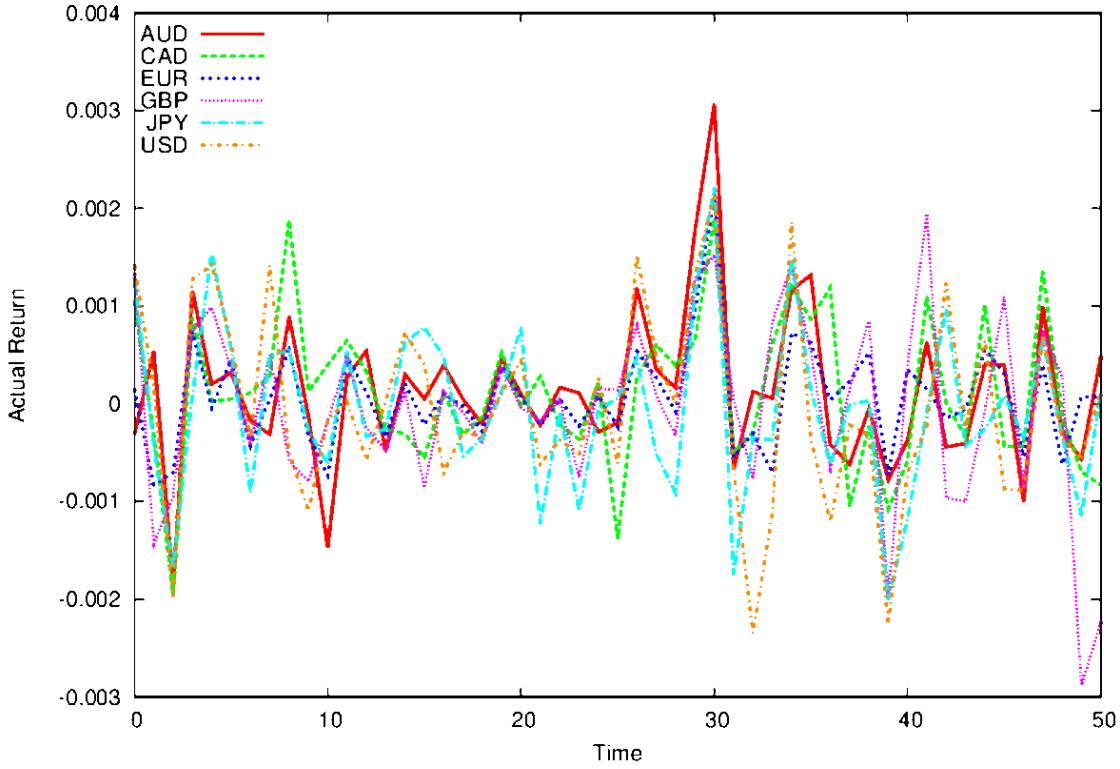
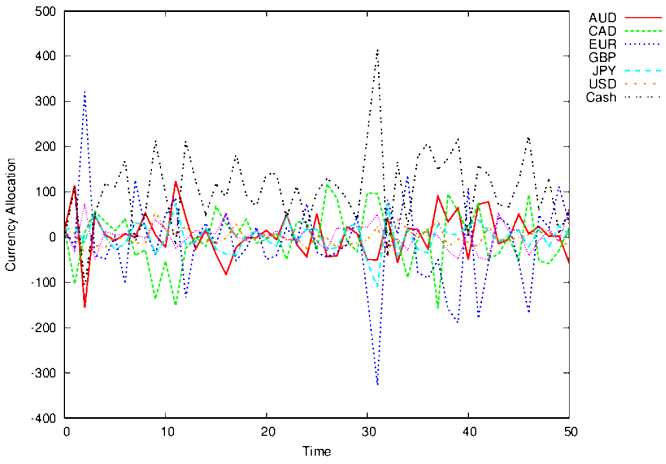
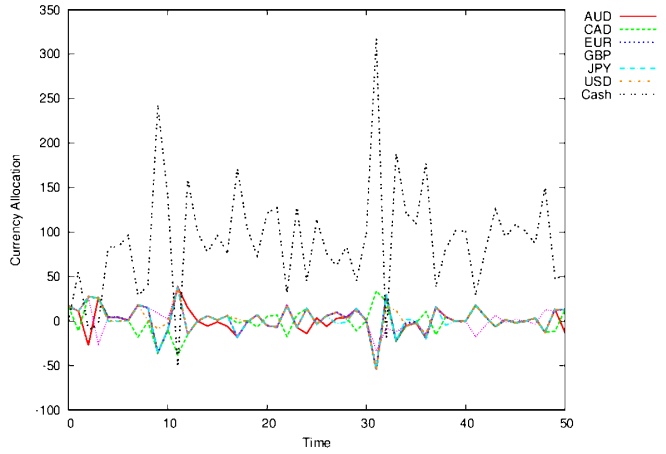


Figure A-3. Realized currency returns on 1/4/2010 from 00:00-12:30
Note: The plots show the realized returns of the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the United States dollar (USD), all in terms of the Swiss franc on 1/4/2010 from 00:00-12:30.



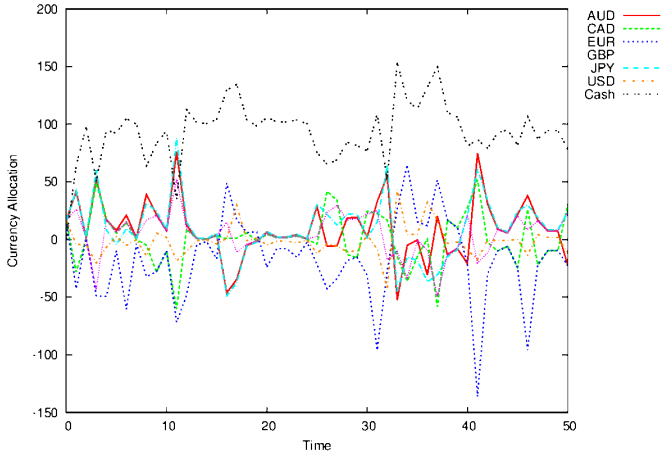
A



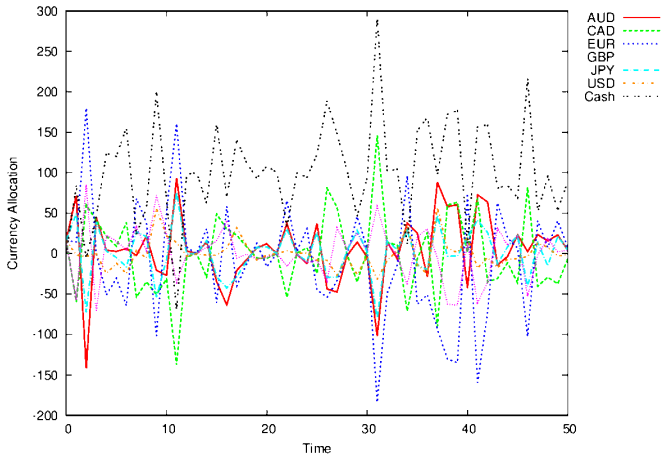
B

Figure A-4. Currency allocation plots on 1/4/2010 from 00:00-12:30

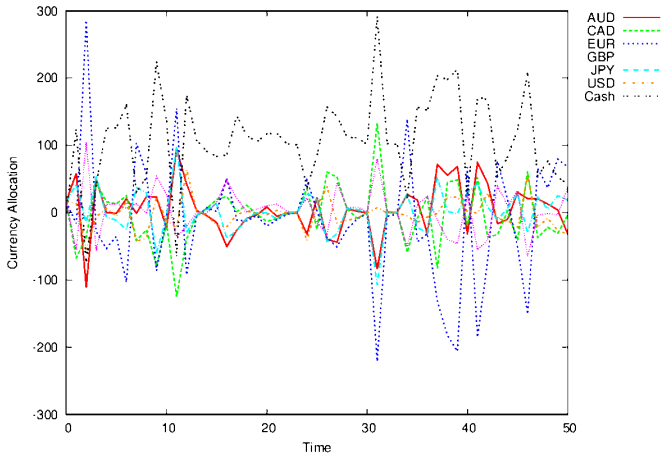
Note: Each plot shows the currency holdings denominated in Swiss francs of a particular portfolio and is labeled according to the constraint set used in its optimization: (A) unconstrained, (B) the relative portfolio weight constraint set (WC), (C) the independent factor relative weight constraint set (WIC), (D) the marginal risk contribution constraint set (MRC), (E) the independent factor marginal risk contribution constraint set (MRIC), and (F) the independent factor total risk contribution constraint set (TRC).



C

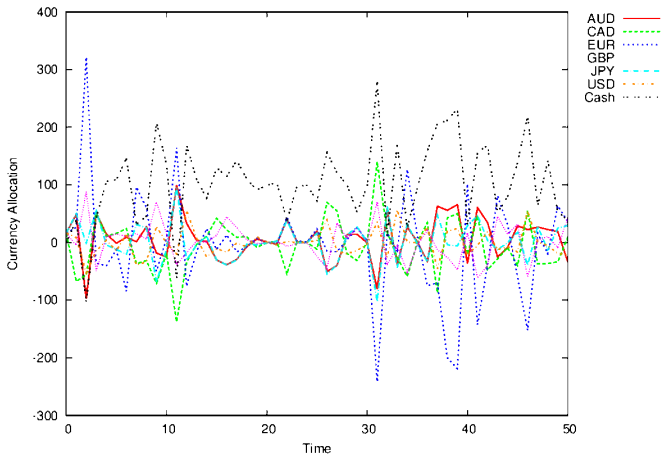


D



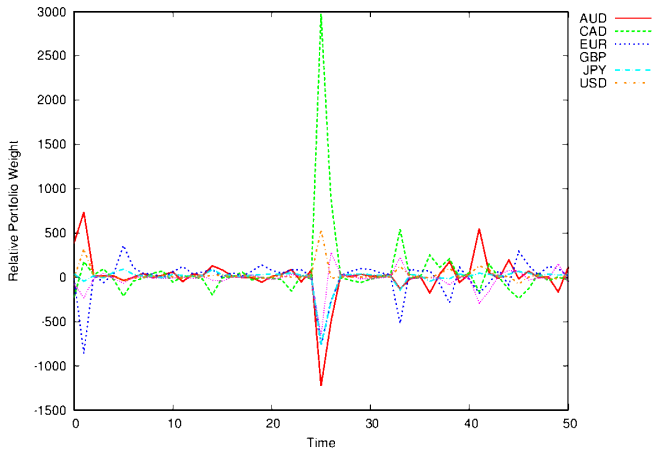
E

Figure A-4 Continued**.**



F

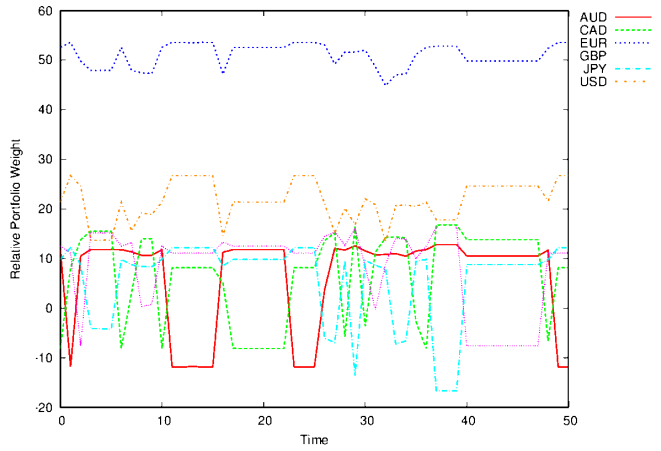
Figure A-4 Continued**.**



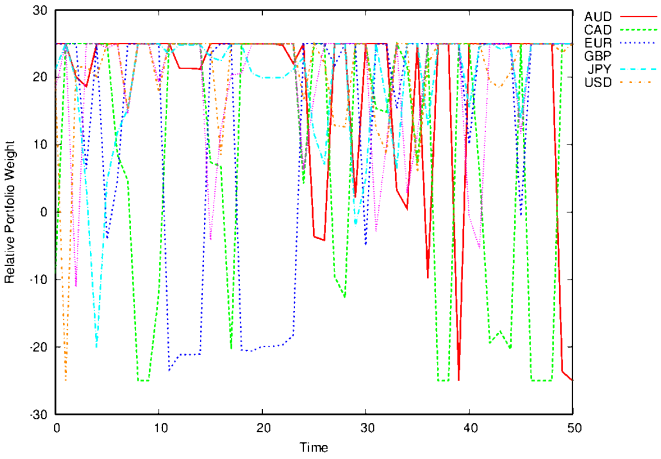
A

Figure A-5. Independent factor relative portfolio weight plots on 1/4/2010 from 00:00-12:30

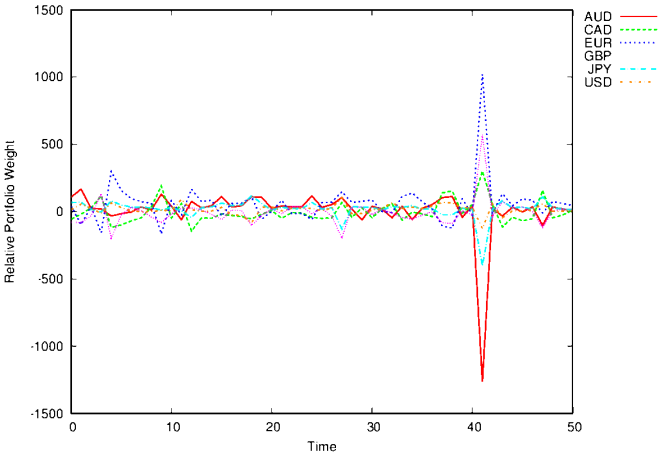
Note: Each plot shows the components of a particular portfolio and is labeled according to the constraint set used in its optimization: (A) unconstrained, (B) the relative portfolio weight constraint set (WC), (C) the independent factor relative weight constraint set (WIC), (D) the marginal risk contribution constraint set (MRC), (E) the independent factor marginal risk contribution constraint set (MRIC), and (F) the independent factor total risk contribution constraint set (TRC).



B

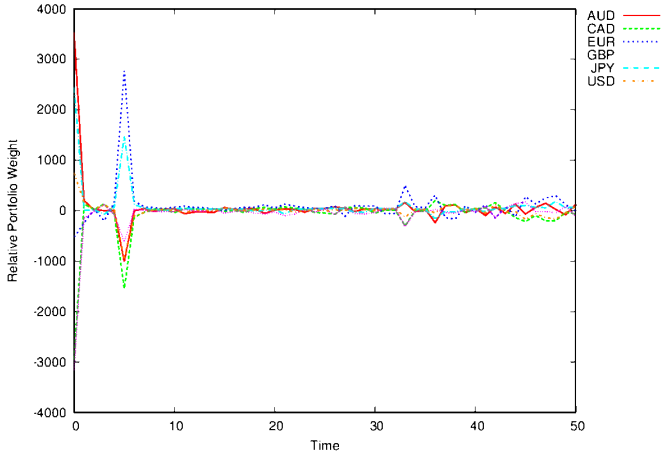


C

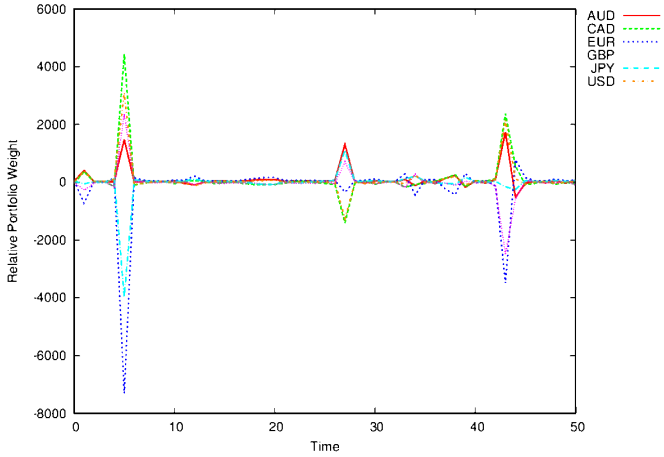


D

Figure A-5 Continued**.**

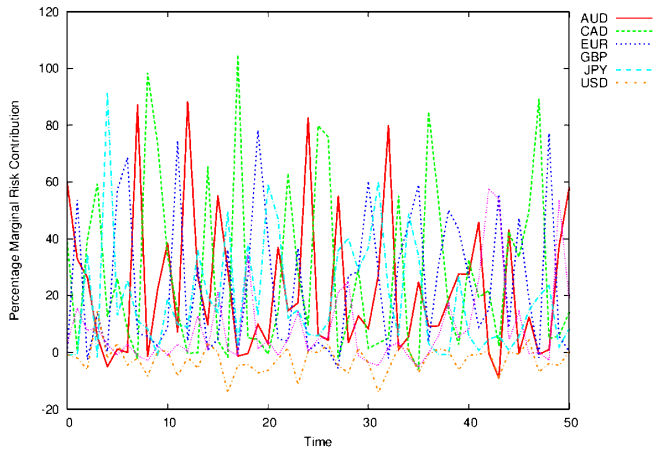


E

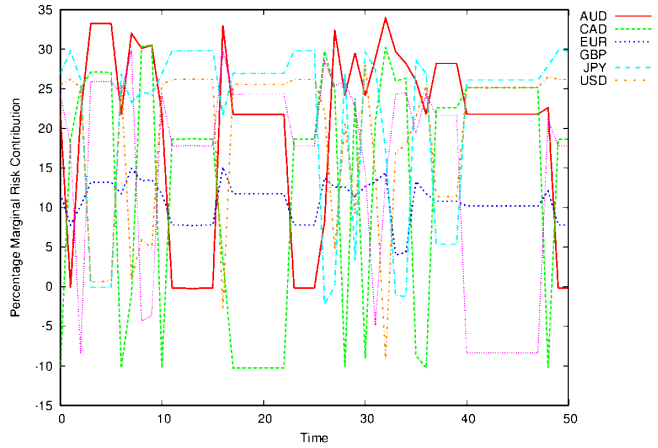


F

Figure A-5 Continued**.**



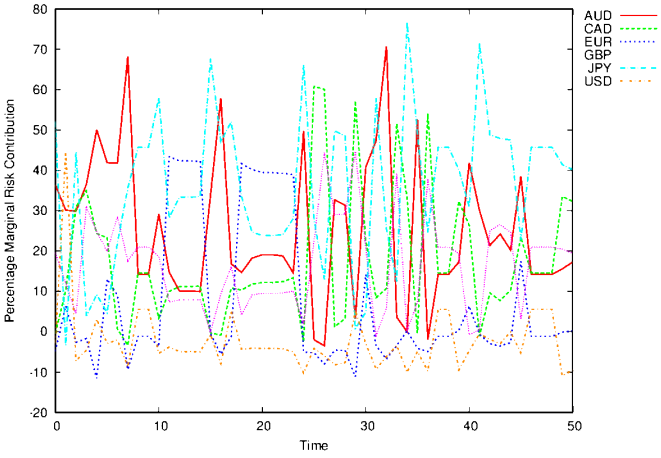
A



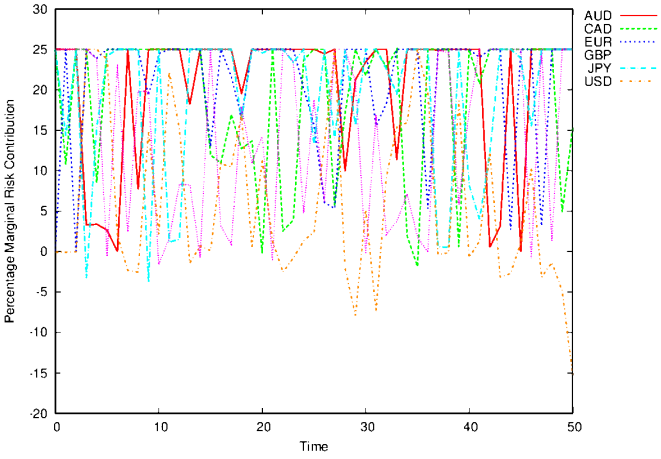
B

Figure A-6. Percentage marginal risk contribution plots on 1/4/2010 from 00:00-12:30

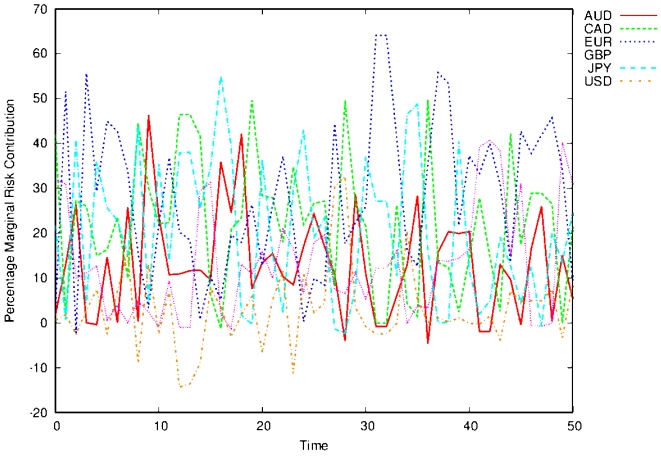
Note: Each plot shows the components of a particular portfolio and is labeled according to the constraint set used in its optimization: (A) unconstrained, (B) the relative portfolio weight constraint set (WC), (C) the independent factor relative weight constraint set (WIC), (D) the marginal risk contribution constraint set (MRC), (E) the independent factor marginal risk contribution constraint set (MRIC), and (F) the independent factor total risk contribution constraint set (TRC).



C

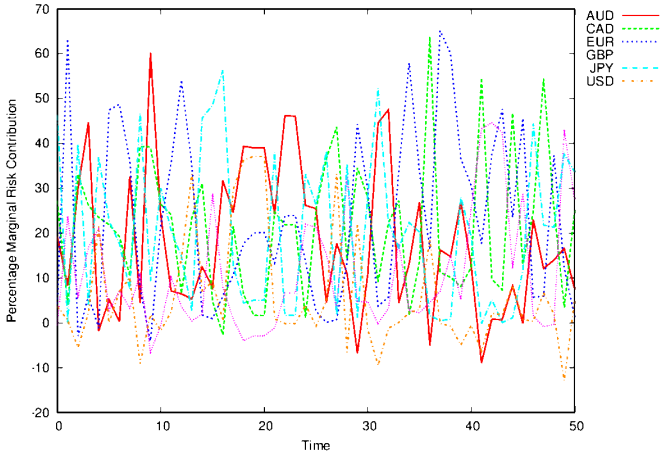


D



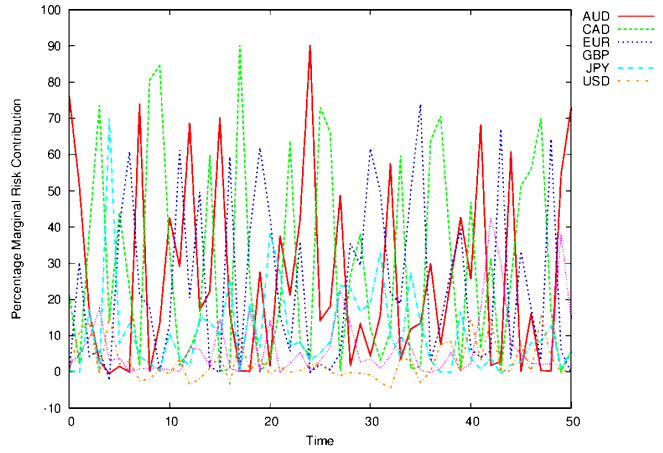
E

Figure A-6 Continued**.**



F

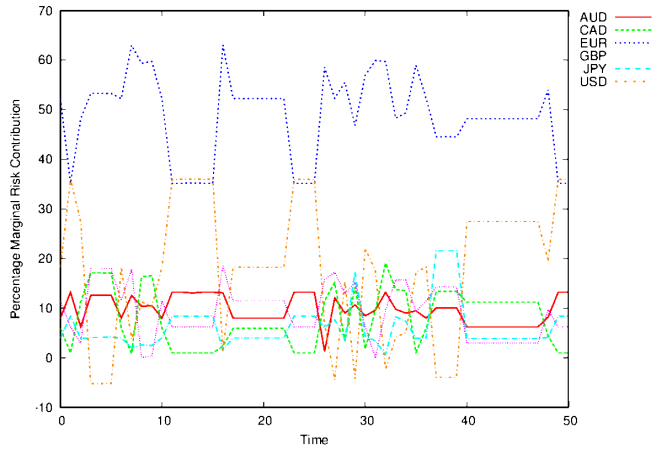
Figure A-6 Continued**.**



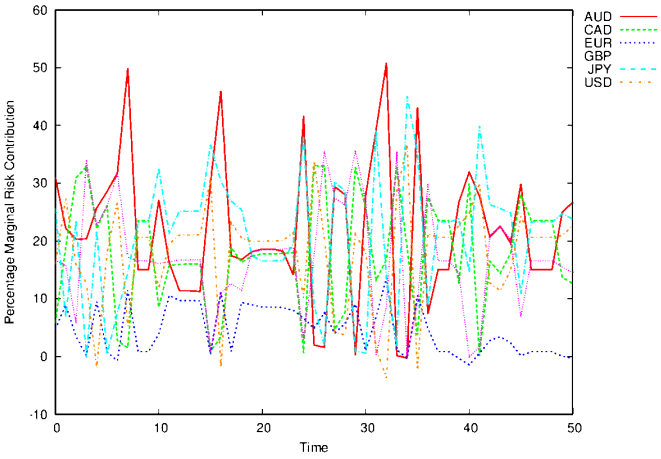
A

Figure A-7. Independent factor percentage marginal risk contribution plots on 1/4/2010 from 00:00-12:30

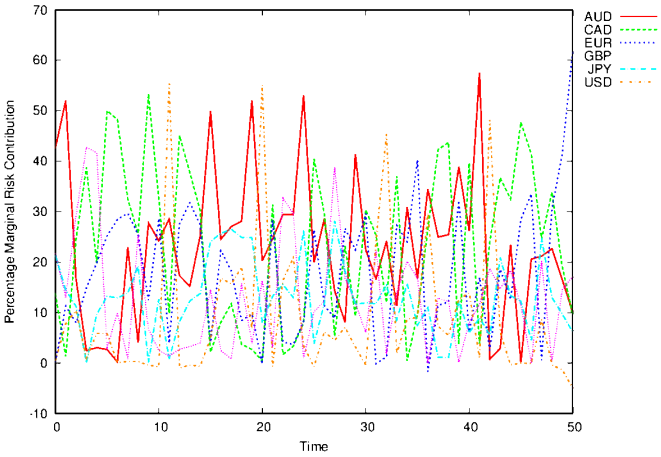
Note: Each plot shows the components of a particular portfolio and is labeled according to the constraint set used in its optimization: (A) unconstrained, (B) the relative portfolio weight constraint set (WC), (C) the independent factor relative weight constraint set (WIC), (D) the marginal risk contribution constraint set (MRC), (E) the independent factor marginal risk contribution constraint set (MRIC), and (F) the independent factor total risk contribution constraint set (TRC).



B

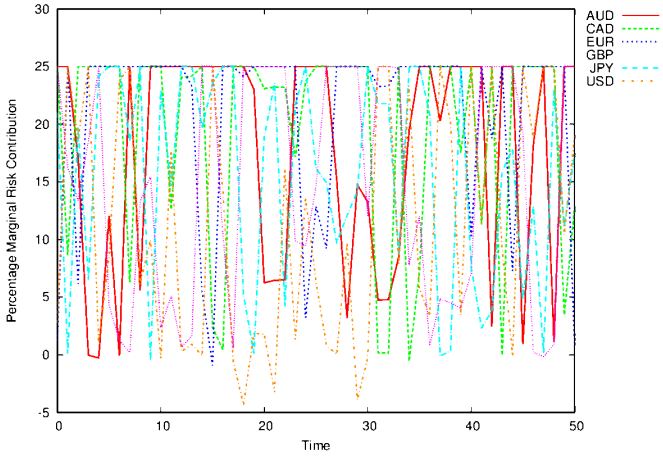


C

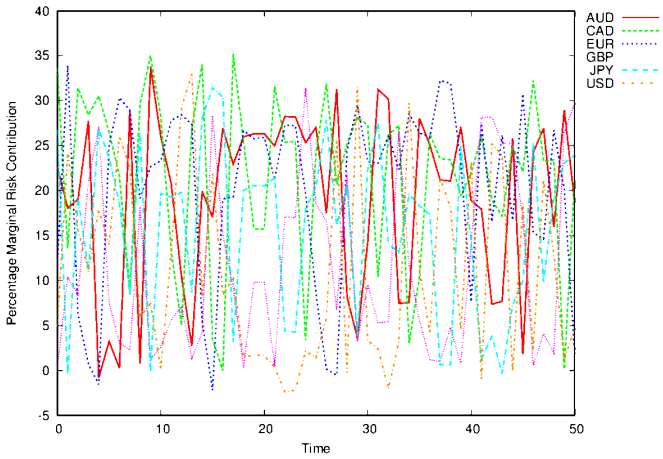


D

Figure A-7 Continued**.**

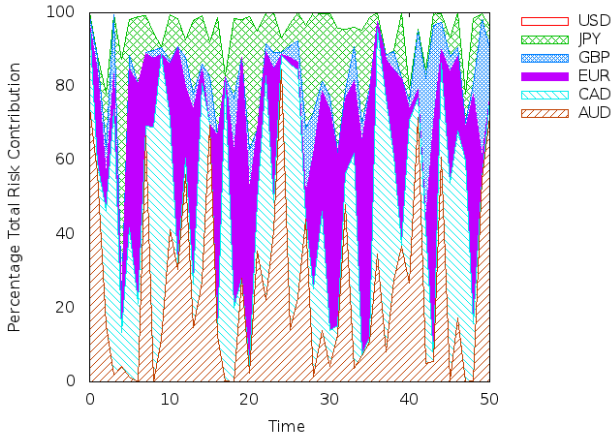


E

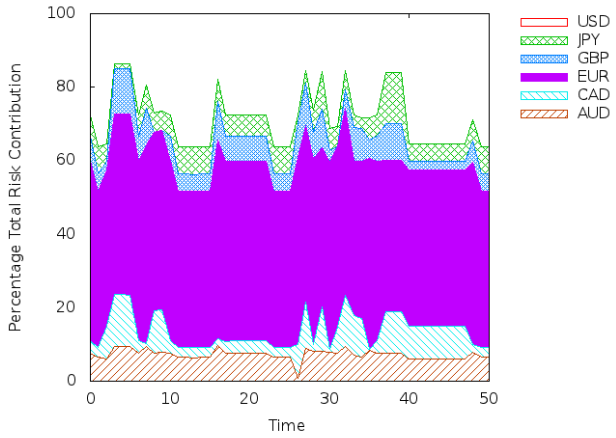


F

Figure A-7 Continued**.**



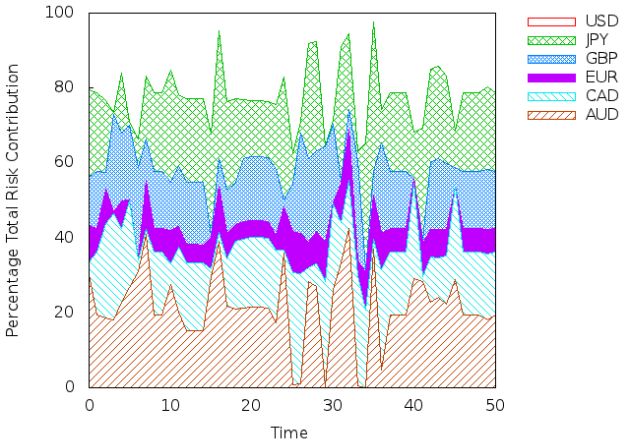
A



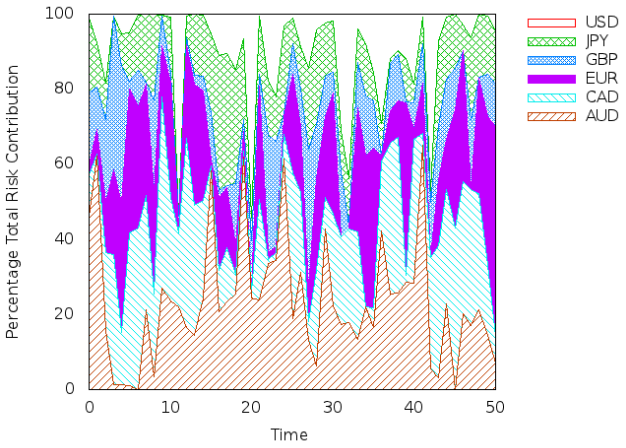
B

Figure A-8. Independent factor total risk contribution plots on 1/4/2010 from 00:00-12:3

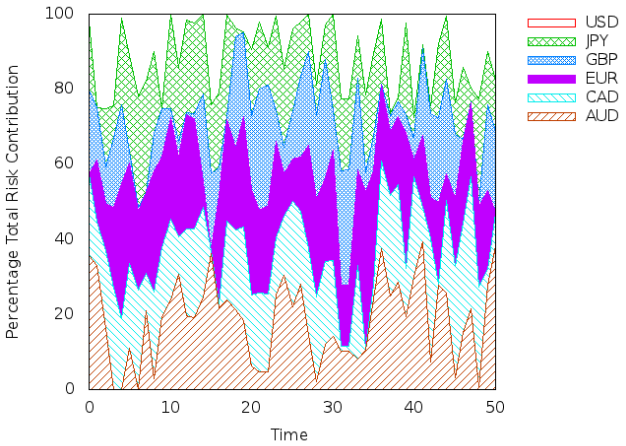
Note: Each plot shows the components of a particular portfolio and is labeled according to the constraint set used in its optimization: (A) unconstrained, (B) the relative portfolio weight constraint set (WC), (C) the independent factor relative weight constraint set (WIC), (D) the marginal risk contribution constraint set (MRC), (E) the independent factor marginal risk contribution constraint set (MRIC), and (F) the independent factor total risk contribution constraint set (TRC).



C

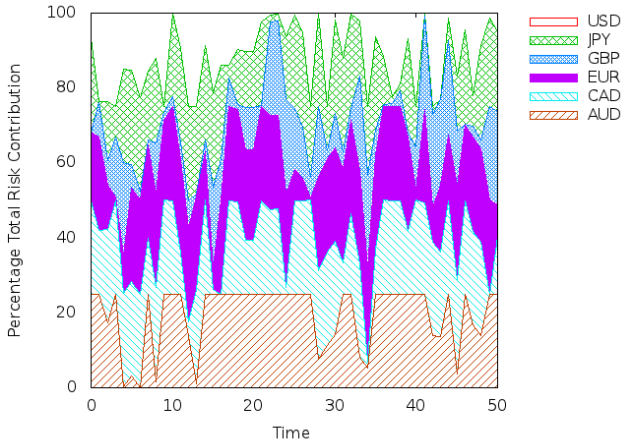


D



E

Figure A-8 Continued.



F

Figure A-8 Continued**.**

Table A-1. Expected Values, 2009

	$\mathbf{E}\{\boldsymbol{R}(t)\}$	$\mathbf{E}\{\hat{\boldsymbol{n}}(t)\}$	$\mathbf{E}\{\hat{\boldsymbol{e}}(t)\}$
AUD	8.614E-06	9.234E-06	7.332E-06
CAD	5.147E-06	5.334E-06	5.639E-06
EUR	-1.644E-07	-3.009E-07	-3.009E-07
GBP	2.639E-06	2.463E-06	2.856E-06
JPY	-2.148E-06	-2.082E-06	-1.189E-06
USD	-1.109E-06	-1.086E-06	-6.905E-07

Note: These are the expected values of the currency log returns $\boldsymbol{R}(t)$, residuals $\hat{\boldsymbol{n}}(t)$ of the VAR-LiNGAM model, and independent factors $\hat{\boldsymbol{e}}(t)$ derived using the VAR-LiNGAM model for the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the United States dollar (USD) in the year of 2009.

Table A-2. Covariance Matrix of the Currency Log Returns, 2009

	AUD	CAD	EUR	GBP	JPY	USD
AUD	1.146E-06	4.832E-07	2.693E-07	4.282E-07	-1.604E-07	1.379E-07
CAD		8.667E-07	2.005E-07	3.476E-07	1.302E-07	3.358E-07
EUR			2.317E-07	2.228E-07	5.855E-08	1.395E-07
GBP				8.172E-07	8.975E-08	2.845E-07
JPY					1.207E-06	5.898E-07
USD						7.528E-07

Note: This is the covariance matrix of the currency log returns $\boldsymbol{R}(t)$ for the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the United States dollar (USD). in the year 2009.

Table A-3. Covariance Matrix of the Residuals, 2009

	AUD	CAD	EUR	GBP	JPY	USD
AUD	1.139E-06	4.841E-07	2.695E-07	4.295E-07	-1.629E-07	1.373E-07
CAD		8.589E-07	2.006E-07	3.475E-07	1.298E-07	3.350E-07
EUR			2.303E-07	2.216E-07	5.716E-08	1.388E-07
GBP				8.157E-07	8.905E-08	2.840E-07
JPY					1.203E-06	5.883E-07
USD						7.520E-07

Note: This is the covariance matrix of the VAR-LiNGAM model residuals $\hat{\boldsymbol{n}}(t)$ for the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the United States dollar (USD) in the year 2009.

Table A-4. Covariance Matrix of Independent Factors, 2009

	AUD	CAD	EUR	GBP	JPY	USD
AUD	7.218E-07	7.370E-08	4.870E-08	6.739E-08	-1.029E-07	-2.164E-07
CAD		6.257E-07	3.093E-08	4.859E-08	-6.907E-08	-1.642E-07
EUR			2.303E-07	4.544E-08	-1.041E-07	-1.642E-07
GBP				5.702E-07	-6.386E-08	-6.532E-08
JPY					7.808E-07	1.245E-07
USD						7.855E-07

Note: This is the covariance matrix of the independent factors $\hat{\boldsymbol{e}}(t)$ derived using the VAR-LiNGAM model for the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the United States dollar (USD) in the year of 2009.

Table A-5. Instantaneous Causal Effect Matrix, 2009

	AUD	CAD	EUR	GBP	JPY	USD
AUD	0.000	0.311	0.460	0.136	-0.136	0.217
CAD	0.000	0.000	0.310	0.110	-0.059	0.557
EUR	0.000	0.000	0.000	0.000	0.000	0.000
GBP	0.000	0.000	0.583	0.000	-0.068	0.331
JPY	0.000	0.000	0.246	0.000	0.000	0.753
USD	0.000	0.000	1.315	0.000	0.000	0.000

Note: This is the VAR-LiNGAM instantaneous causal effect matrix $\boldsymbol{B}_0$ for the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the United States dollar (USD) in the year 2009.

Table A-6. Lag One Causal Effect Matrix, 2009

	AUD	CAD	EUR	GBP	JPY	USD
AUD	-0.113	0.053	0.068	0.044	-0.019	0.004
CAD	0.030	-0.106	0.079	0.028	-0.018	0.015
EUR	0.011	0.006	-0.080	-0.006	-0.021	0.028
GBP	0.017	-0.001	-0.001	-0.025	-0.003	0.020
JPY	-0.028	0.018	-0.005	0.009	-0.032	0.006
USD	-0.018	-0.009	0.075	0.008	0.002	-0.024

Note: This is the VAR-LiNGAM lag one causal effect matrix $\boldsymbol{B}_1$ for the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the United States dollar (USD) in the year 2009.

Table A-7. Lag One Autoregressive Matrix, 2009

	AUD	CAD	EUR	GBP	JPY	USD
AUD	-0.091	0.020	0.034	0.044	-0.040	0.032
CAD	0.036	-0.106	0.034	0.024	-0.037	0.033
EUR	0.011	0.006	-0.080	-0.006	-0.021	0.028
GBP	0.024	0.000	-0.055	-0.029	-0.020	0.038
JPY	-0.028	0.018	-0.049	0.008	-0.056	0.022
USD	-0.003	-0.001	-0.031	0.001	-0.025	0.012

Note: This is the VAR-LiNGAM lag one matrix of autoregressive least-squares estimates $\boldsymbol{M_1}$ for the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the United States dollar (USD) in the year 2009.

APPENDIX B

CHAPTER III TABLES

Table B-1. Expected Values of the Log Returns, 2009-2011

	2009	2010	2011
AUD	9.592E-06	1.026E-06	2.060E-07
CAD	5.043E-06	-2.211E-06	-9.024E-07
EUR	1.545E-07	-7.262E-06	-1.032E-06
GBP	3.540E-06	-5.755E-06	1.706E-07
JPY	-1.999E-06	1.403E-06	2.471E-06
USD	-1.096E-06	-4.330E-06	2.143E-07

Note: These are the expected values of the currency log returns $\boldsymbol{R}(t)$ in the years 2009-2011 for the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the United States dollar (USD).

Table B-2. Covariance Matrices of the Log Returns, 2009-2011

A

	AUD	CAD	EUR	GBP	JPY	USD
AUD	1.256E-06	5.408E-07	3.160E-07	4.863E-07	-1.598E-07	1.570E-07
CAD	5.408E-07	9.250E-07	2.313E-07	3.882E-07	1.319E-07	3.508E-07
EUR	3.160E-07	2.313E-07	2.752E-07	2.635E-07	7.758E-08	1.663E-07
GBP	4.863E-07	3.882E-07	2.635E-07	8.979E-07	1.112E-07	3.170E-07
JPY	-1.598E-07	1.319E-07	7.758E-08	1.112E-07	1.286E-06	6.305E-07
USD	1.570E-07	3.508E-07	1.663E-07	3.170E-07	6.305E-07	7.994E-07

B

	AUD	CAD	EUR	GBP	JPY	USD
AUD	9.002E-07	5.584E-07	3.692E-07	3.882E-07	7.841E-08	2.767E-07
CAD	5.584E-07	7.459E-07	3.172E-07	3.733E-07	1.782E-07	3.672E-07
EUR	3.692E-07	3.172E-07	3.848E-07	2.903E-07	1.328E-07	2.041E-07
GBP	3.882E-07	3.733E-07	2.903E-07	6.068E-07	2.471E-07	3.375E-07
JPY	7.841E-08	1.782E-07	1.328E-07	2.471E-07	1.020E-06	5.344E-07
USD	2.767E-07	3.672E-07	2.041E-07	3.375E-07	5.344E-07	6.085E-07

Table B-2 Continued**.**

C

	AUD	CAD	EUR	GBP	JPY	USD
AUD	1.516E-06	1.148E-06	1.027E-06	9.714E-07	6.579E-07	8.386E-07
CAD	1.148E-06	1.267E-06	9.301E-07	9.454E-07	7.144E-07	9.273E-07
EUR	1.027E-06	9.301E-07	1.048E-06	8.841E-07	6.242E-07	7.573E-07
GBP	9.714E-07	9.454E-07	8.841E-07	1.083E-06	7.254E-07	8.861E-07
JPY	6.579E-07	7.144E-07	6.242E-07	7.254E-07	1.102E-06	8.525E-07
USD	8.386E-07	9.273E-07	7.573E-07	8.861E-07	8.525E-07	1.071E-06

Note: These are the covariance matrices of the currency log returns $\boldsymbol{R}(t)$ in the years 2009 (A), 2010 (B), and 2011 (C) for the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the United States dollar (USD).

Table B-3. Expected Values of the Residuals, **2009-2011**

	2009	**2010**	**2011**
AUD	1.033E-05	1.146E-06	2.872E-07
CAD	5.268E-06	-2.256E-06	-8.374E-07
EUR	7.928E-10	-7.575E-06	-9.918E-07
GBP	3.459E-06	-6.008E-06	2.397E-07
JPY	-1.950E-06	1.330E-06	2.623E-06
USD	-1.155E-06	-4.444E-06	2.858E-07

Note: These are the expected values of the estimated vector autoregression residuals $\widehat{\boldsymbol{n}}(t)$ in the years 2009-2011 for the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the United States dollar (USD).

Table B-4. Covariance Matrix of the Residuals, 2009-2011

A

	AUD	CAD	EUR	GBP	JPY	USD
AUD	1.245E-06	5.414E-07	3.150E-07	4.871E-07	-1.637E-07	1.555E-07
CAD	5.414E-07	9.159E-07	2.310E-07	3.879E-07	1.299E-07	3.496E-07
EUR	3.150E-07	2.310E-07	2.716E-07	2.605E-07	7.333E-08	1.635E-07
GBP	4.871E-07	3.879E-07	2.605E-07	8.944E-07	1.078E-07	3.147E-07
JPY	-1.637E-07	1.299E-07	7.333E-08	1.078E-07	1.279E-06	6.263E-07
USD	1.555E-07	3.496E-07	1.635E-07	3.147E-07	6.263E-07	7.968E-07

Table B-4 Continued**.**

B

	AUD	CAD	EUR	GBP	JPY	USD
AUD	8.971E-07	5.570E-07	3.673E-07	3.863E-07	7.600E-08	2.746E-07
CAD	5.570E-07	7.433E-07	3.154E-07	3.717E-07	1.756E-07	3.657E-07
EUR	3.673E-07	3.154E-07	3.828E-07	2.885E-07	1.305E-07	2.024E-07
GBP	3.863E-07	3.717E-07	2.885E-07	6.046E-07	2.450E-07	3.359E-07
JPY	7.600E-08	1.756E-07	1.305E-07	2.450E-07	1.017E-06	5.323E-07
USD	2.746E-07	3.657E-07	2.024E-07	3.359E-07	5.323E-07	6.068E-07

C

	AUD	CAD	EUR	GBP	JPY	USD
AUD	1.514E-06	1.148E-06	1.026E-06	9.704E-07	6.576E-07	8.379E-07
CAD	1.148E-06	1.265E-06	9.285E-07	9.440E-07	7.133E-07	9.262E-07
EUR	1.026E-06	9.285E-07	1.046E-06	8.826E-07	6.230E-07	7.561E-07
GBP	9.704E-07	9.440E-07	8.826E-07	1.081E-06	7.245E-07	8.851E-07
JPY	6.576E-07	7.133E-07	6.230E-07	7.245E-07	1.101E-06	8.516E-07
USD	8.379E-07	9.262E-07	7.561E-07	8.851E-07	8.516E-07	1.070E-06

Note: These are the covariance matrices of the estimated vector autoregression residuals $\hat{\boldsymbol{n}}(t)$ in the years 2009 (A), 2010 (B), and 2011 (C) for the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the United States dollar (USD).

Table B-5. Contemporaneous Causal Effect Matrices, 2009-2011

A

	AUD	CAD	EUR	GBP	JPY	USD
AUD	0.000	0.284	0.470	0.145	0.000	0.203
CAD	0.000	0.000	0.289	0.102	0.000	0.527
EUR	0.000	0.000	0.000	0.000	0.000	0.000
GBP	0.000	0.000	0.579	0.000	0.000	0.332
JPY	-0.146	-0.041	0.297	-0.022	0.000	0.752
USD	0.000	0.000	1.226	0.000	0.000	0.000

Table B-5 Continued**.**

B

	AUD	CAD	EUR	GBP	JPY	USD
AUD	0.000	0.000	0.452	0.000	0.000	0.000
CAD	0.237	0.000	0.256	0.064	-0.083	0.530
EUR	0.000	0.000	0.000	0.000	0.000	0.000
GBP	0.086	0.000	0.396	0.000	0.017	0.426
JPY	-0.164	0.000	0.072	0.000	0.000	1.086
USD	-0.297	0.000	1.614	0.000	0.000	0.000

C

	AUD	CAD	EUR	GBP	JPY	USD
AUD	0.000	0.000	0.407	0.000	0.000	0.000
CAD	0.241	0.000	0.198	0.066	0.000	0.530
EUR	0.000	0.000	0.000	0.000	0.000	0.000
GBP	0.074	0.000	0.376	0.000	0.000	0.495
JPY	-0.001	-0.071	0.054	0.069	0.000	0.863
USD	-0.387	0.000	1.703	0.000	0.000	0.000

Note: These are the VAR-LiNGAM model estimates of the causal effect matrices $\mathbf{B}_0$ in the years 2009 (A), 2010 (B), and 2011 (C) for the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the United States dollar (USD). A $\mathbf{B}_0$ matrix contains the causal affects within the lag 0 period. For example in 2009, the effect on the AUD's log return in the lag 0 period from the CAD's log return in the lag 0 period is 0.28387 times the value of the CAD's lag 0 log return.

Table B-6. Lag One causal Effect Matrices, 2009-2011

A

	AUD	CAD	EUR	GBP	JPY	USD
AUD	-0.132	0.057	0.112	0.051	-0.019	-0.002
CAD	0.025	-0.118	0.088	0.033	-0.019	0.019
EUR	0.007	0.011	-0.129	-0.001	-0.023	0.036
GBP	0.018	-0.004	0.012	-0.037	-0.001	0.024
JPY	-0.036	0.008	-0.008	0.008	-0.048	0.024
USD	-0.011	-0.010	0.075	0.000	-0.003	-0.023

Table B-6 Continued**.**

B

	AUD	CAD	EUR	GBP	JPY	USD
AUD	-0.056	0.031	0.020	-0.001	-0.023	-0.004
CAD	0.038	-0.044	-0.018	0.013	-0.027	0.043
EUR	-0.005	0.000	-0.047	-0.006	-0.028	0.011
GBP	0.012	-0.007	0.022	-0.037	0.002	0.002
JPY	-0.002	-0.015	-0.008	0.003	-0.030	0.036
USD	-0.024	0.027	0.053	0.003	0.003	-0.022

C

	AUD	CAD	EUR	GBP	JPY	USD
AUD	-0.038	0.070	-0.015	-0.009	0.012	-0.026
CAD	0.031	-0.042	0.008	-0.019	-0.016	0.037
EUR	0.022	0.027	-0.071	0.000	-0.033	0.027
GBP	0.004	-0.001	0.026	-0.045	-0.005	0.010
JPY	0.013	-0.009	0.010	0.011	-0.019	0.000
USD	-0.042	0.004	0.073	-0.004	0.021	-0.025

Note: These are the VAR-LiNGAM model estimates of the causal effect matrices $\mathbf{B}_1$ in the years 2009 (A), 2010 (B), and 2011 (C) for the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the United States dollar (USD). A $\mathbf{B}_1$ matrix contains the causal affects from the lag 1 period on the lag 0 period. For example in 2009, the effect on the AUD's log return in the lag 0 period from the CAD's log return in the lag 1 period is 0.056585 times the value of the CAD's lag 1 log return.

Table B-7. Autoregressive matrices, 2009-2011

A

	AUD	CAD	EUR	GBP	JPY	USD
AUD	-0.119	0.031	0.021	0.052	-0.052	0.039
CAD	0.028	-0.113	-0.002	0.028	-0.045	0.046
EUR	0.007	0.011	-0.129	-0.001	-0.023	0.036
GBP	0.021	0.003	-0.090	-0.038	-0.025	0.052
JPY	-0.020	0.014	-0.110	-0.001	-0.068	0.041
USD	-0.003	0.004	-0.083	-0.001	-0.031	0.020

Table B-7 Continued**.**

B

	AUD	CAD	EUR	GBP	JPY	USD
AUD	-0.058	0.031	-0.001	-0.004	-0.036	0.001
CAD	0.016	-0.027	-0.040	0.006	-0.056	0.041
EUR	-0.005	0.000	-0.047	-0.006	-0.028	0.011
GBP	-0.001	0.004	-0.007	-0.041	-0.027	0.005
JPY	-0.009	-0.001	-0.035	-0.002	-0.060	0.033
USD	-0.015	0.018	-0.023	-0.005	-0.031	-0.004

C

	AUD	CAD	EUR	GBP	JPY	USD
AUD	-0.029	0.081	-0.044	-0.009	-0.001	-0.015
CAD	0.032	-0.005	-0.035	-0.024	-0.043	0.055
EUR	0.022	0.027	-0.071	0.000	-0.033	0.027
GBP	0.013	0.024	-0.019	-0.046	-0.034	0.032
JPY	0.018	0.011	-0.019	0.009	-0.049	0.023
USD	0.005	0.019	-0.031	-0.001	-0.034	0.026

Note: These are the VAR-LiNGAM model estimates of the autoregressive matrices $\mathbf{M}_1$ in the years 2009 (A), 2010 (B), and 2011 (C) for the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the United States dollar (USD). These matrices contain the autoregressive affects from the lag 1 period on the lag 0 period. An $\mathbf{M}_1$ matrix contains the estimates from a standard vector autoregressive model. An $\mathbf{M}_1$ matrix is modified by the results of the LiNGAM algorithm to produce a $\mathbf{B}_1$ causal effect matrix.

APPENDIX C

CHAPTER IV TABLES

Table C-1. Expected Values of the Log Returns in the Estimation and Forecast Data Sets

A

	Before	**Surrounding**	**After**	**Long After**
AUD	-7.152E-06	-2.378E-05	-5.459E-06	5.858E-06
CAD	-1.028E-05	-2.471E-05	-6.673E-06	4.508E-06
EUR	-6.023E-06	-2.153E-05	-5.508E-06	-1.065E-07
GBP	-1.068E-05	-2.472E-05	-4.399E-06	4.276E-06
JPY	-9.118E-06	-1.776E-05	8.366E-06	2.120E-06
USD	-1.288E-05	-2.483E-05	-5.025E-07	5.555E-06

B

	Before	**Surrounding**	**After**	**Long After**
AUD	-4.603E-05	5.066E-05	1.343E-05	-9.477E-06
CAD	-4.201E-05	4.705E-05	8.368E-06	2.832E-06
EUR	-4.136E-05	4.640E-05	1.111E-06	-9.271E-07
GBP	-4.005E-05	4.833E-05	6.015E-06	8.458E-06
JPY	-2.687E-05	5.640E-05	2.508E-06	4.670E-06
USD	-3.853E-05	5.600E-05	5.585E-06	2.018E-06

Note: These are the expected values of the currency log returns $\boldsymbol{R}(t)$ in the estimation data sets (A) and the forecast data sets (B) for the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the United States dollar (USD).

Table C-2. Estimation Data Set Covariance Matrices

A

	AUD	CAD	EUR	GBP	JPY	USD
AUD	8.451E-07	5.718E-07	4.437E-07	4.697E-07	2.048E-07	4.125E-07
CAD	5.718E-07	7.161E-07	3.834E-07	4.532E-07	2.469E-07	4.739E-07
EUR	4.437E-07	3.834E-07	4.812E-07	3.709E-07	1.626E-07	2.842E-07
GBP	4.697E-07	4.532E-07	3.709E-07	6.026E-07	2.483E-07	4.116E-07
JPY	2.048E-07	2.469E-07	1.626E-07	2.483E-07	6.497E-07	3.348E-07
USD	4.125E-07	4.739E-07	2.842E-07	4.116E-07	3.348E-07	5.612E-07

B

	AUD	CAD	EUR	GBP	JPY	USD
AUD	1.363E-06	9.944E-07	8.574E-07	7.918E-07	3.853E-07	6.824E-07
CAD	9.944E-07	1.090E-06	7.452E-07	7.359E-07	4.216E-07	7.104E-07
EUR	8.574E-07	7.452E-07	8.543E-07	6.639E-07	3.380E-07	5.260E-07
GBP	7.918E-07	7.359E-07	6.639E-07	8.203E-07	3.939E-07	6.071E-07
JPY	3.853E-07	4.216E-07	3.380E-07	3.939E-07	8.074E-07	4.824E-07
USD	6.824E-07	7.104E-07	5.260E-07	6.071E-07	4.824E-07	7.341E-07

C

	AUD	CAD	EUR	GBP	JPY	USD
AUD	2.156E-06	1.742E-06	1.647E-06	1.507E-06	1.108E-06	1.334E-06
CAD	1.742E-06	1.841E-06	1.511E-06	1.458E-06	1.142E-06	1.401E-06
EUR	1.647E-06	1.511E-06	1.672E-06	1.422E-06	1.054E-06	1.242E-06
GBP	1.507E-06	1.458E-06	1.422E-06	1.566E-06	1.141E-06	1.332E-06
JPY	1.108E-06	1.142E-06	1.054E-06	1.141E-06	1.417E-06	1.255E-06
USD	1.334E-06	1.401E-06	1.242E-06	1.332E-06	1.255E-06	1.489E-06

D

	AUD	CAD	EUR	GBP	JPY	USD
AUD	6.878E-07	4.024E-07	2.445E-07	2.584E-07	1.824E-07	1.782E-07
CAD	4.024E-07	5.528E-07	2.151E-07	3.146E-07	3.239E-07	3.685E-07
EUR	2.445E-07	2.151E-07	2.510E-07	1.980E-07	1.549E-07	1.641E-07
GBP	2.584E-07	3.146E-07	1.980E-07	4.739E-07	3.710E-07	4.106E-07
JPY	1.824E-07	3.239E-07	1.549E-07	3.710E-07	8.613E-07	6.297E-07
USD	1.782E-07	3.685E-07	1.641E-07	4.106E-07	6.297E-07	7.180E-07

Note: These are the covariance matrices of the currency log returns $\boldsymbol{R}(t)$ in the before (A), surrounding (B), after (C), and long after (D) estimation data sets for the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the United States dollar (USD).

Table C-3. Forecast Data Set Covariance Matrices

A

	AUD	CAD	EUR	GBP	JPY	USD
AUD	2.211E-06	1.747E-06	1.621E-06	1.394E-06	9.397E-07	1.205E-06
CAD	1.747E-06	1.799E-06	1.433E-06	1.292E-06	8.919E-07	1.200E-06
EUR	1.621E-06	1.433E-06	1.612E-06	1.262E-06	8.046E-07	1.017E-06
GBP	1.394E-06	1.292E-06	1.262E-06	1.306E-06	8.113E-07	1.026E-06
JPY	9.397E-07	8.919E-07	8.046E-07	8.113E-07	1.022E-06	8.514E-07
USD	1.205E-06	1.200E-06	1.017E-06	1.026E-06	8.514E-07	1.127E-06

B

	AUD	CAD	EUR	GBP	JPY	USD
AUD	3.475E-06	2.948E-06	2.912E-06	2.745E-06	2.249E-06	2.479E-06
CAD	2.948E-06	3.039E-06	2.746E-06	2.692E-06	2.338E-06	2.601E-06
EUR	2.912E-06	2.746E-06	2.979E-06	2.718E-06	2.273E-06	2.493E-06
GBP	2.745E-06	2.692E-06	2.718E-06	2.901E-06	2.395E-06	2.623E-06
JPY	2.249E-06	2.338E-06	2.273E-06	2.395E-06	2.576E-06	2.539E-06
USD	2.479E-06	2.601E-06	2.493E-06	2.623E-06	2.539E-06	2.814E-06

C

	AUD	CAD	EUR	GBP	JPY	USD
AUD	8.592E-07	5.108E-07	3.505E-07	3.336E-07	2.057E-07	1.848E-07
CAD	5.108E-07	6.702E-07	3.019E-07	3.807E-07	3.474E-07	3.996E-07
EUR	3.505E-07	3.019E-07	3.407E-07	2.759E-07	2.142E-07	2.224E-07
GBP	3.336E-07	3.807E-07	2.759E-07	5.475E-07	4.203E-07	4.713E-07
JPY	2.057E-07	3.474E-07	2.142E-07	4.203E-07	1.087E-06	7.074E-07
USD	1.848E-07	3.996E-07	2.224E-07	4.713E-07	7.074E-07	8.208E-07

D

	AUD	CAD	EUR	GBP	JPY	USD
AUD	3.302E-07	1.668E-07	2.657E-08	8.621E-08	4.064E-08	9.381E-08
CAD	1.668E-07	2.756E-07	2.446E-08	1.152E-07	1.016E-07	1.862E-07
EUR	2.657E-08	2.446E-08	2.873E-08	2.303E-08	1.260E-08	2.271E-08
GBP	8.621E-08	1.152E-07	2.303E-08	1.886E-07	1.228E-07	1.492E-07
JPY	4.064E-08	1.016E-07	1.260E-08	1.228E-07	6.001E-07	2.567E-07
USD	9.381E-08	1.862E-07	2.271E-08	1.492E-07	2.567E-07	2.977E-07

Note: These are the covariance matrices of the currency log returns $\boldsymbol{R}(t)$ in the before (A), surrounding (B), after (C), and long after (D) forecast data sets for the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the United States dollar (USD).

Table C-4. Estimation Data Set Autoregressive Matrices

A

	AUD	CAD	EUR	GBP	JPY	USD
AUD	-0.028	0.060	-0.047	0.023	0.001	-0.050
CAD	0.021	-0.014	-0.036	0.015	-0.042	0.026
EUR	0.028	0.009	-0.064	0.026	-0.044	0.001
GBP	0.009	0.008	-0.025	0.000	-0.032	-0.010
JPY	0.016	-0.026	-0.026	0.043	-0.072	-0.009
USD	0.001	0.010	-0.037	0.027	-0.025	-0.018

B

	AUD	CAD	EUR	GBP	JPY	USD
AUD	-0.015	0.111	-0.066	-0.031	-0.012	0.012
CAD	0.036	0.028	-0.047	-0.038	-0.056	0.069
EUR	0.030	0.061	-0.084	-0.016	-0.046	0.023
GBP	0.021	0.052	-0.033	-0.061	-0.040	0.033
JPY	0.024	0.028	-0.028	0.017	-0.084	0.018
USD	0.023	0.035	-0.057	0.004	-0.038	0.013

C

	AUD	CAD	EUR	GBP	JPY	USD
AUD	0.005	0.090	-0.068	-0.028	-0.010	0.002
CAD	0.045	0.004	-0.041	-0.044	-0.054	0.078
EUR	0.035	0.036	-0.079	-0.017	-0.023	0.027
GBP	0.030	0.028	-0.023	-0.075	-0.028	0.051
JPY	0.021	0.014	-0.003	0.001	-0.058	0.030
USD	0.024	0.006	-0.030	-0.002	-0.032	0.032

D

	AUD	CAD	EUR	GBP	JPY	USD
AUD	-0.050	0.064	-0.101	-0.008	0.029	-0.047
CAD	0.037	-0.041	-0.103	-0.033	-0.002	0.025
EUR	0.004	0.005	-0.149	0.004	0.005	0.000
GBP	0.002	0.017	-0.116	-0.050	-0.008	0.020
JPY	0.005	0.017	-0.134	-0.018	0.017	-0.009
USD	-0.005	0.013	-0.112	-0.023	-0.016	0.023

Note: These are the VAR and VAR-LiNGAM model estimates of the autoregressive matrices $\mathbf{M_1}$ in the before (A), surrounding (B), after (C), and long after (D) estimation data sets for the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the United States dollar (USD). An $\mathbf{M_1}$ matrix contains the estimates from a standard vector autoregressive model and reflects the autoregressive affects from the lag 1 period on the lag 0 period.

Table C-5. Estimation Data Set Contemporaneous Causal Effect Matrices

A

	AUD	CAD	EUR	GBP	JPY	USD
AUD	0.000	0.218	0.327	0.087	-0.090	0.432
CAD	0.000	0.000	0.111	0.064	0.000	0.610
EUR	0.000	0.000	0.000	0.000	0.000	0.000
GBP	0.000	0.000	0.314	0.000	0.000	0.493
JPY	0.000	-0.101	0.027	0.053	0.000	0.861
USD	0.000	0.000	-1.397	0.000	0.000	0.000

B

	AUD	CAD	EUR	GBP	JPY	USD
AUD	0.000	0.285	0.373	0.096	-0.084	0.381
CAD	0.000	0.000	0.209	0.090	0.000	0.580
EUR	0.000	0.000	0.000	0.000	0.000	0.000
GBP	0.000	0.000	0.339	0.000	0.000	0.535
JPY	0.000	-0.115	0.018	0.050	0.000	0.917
USD	0.000	0.000	-1.285	0.000	0.000	0.000

C

	AUD	CAD	EUR	GBP	JPY	USD
AUD	0.000	0.000	0.390	0.000	0.000	0.000
CAD	0.243	0.000	0.209	0.082	-0.063	0.535
EUR	0.000	0.000	0.000	0.000	0.000	0.000
GBP	0.063	0.000	0.370	0.000	0.000	0.522
JPY	-0.024	0.000	0.058	0.064	0.000	0.866
USD	-0.405	0.000	1.604	0.000	0.000	0.000

D

	AUD	CAD	EUR	GBP	JPY	USD
AUD	0.000	0.000	0.446	0.000	0.000	0.000
CAD	0.256	0.000	0.220	0.057	-0.020	0.476
EUR	0.000	0.000	0.000	0.000	0.000	0.000
GBP	0.081	0.000	0.369	0.000	0.000	0.485
JPY	0.002	0.000	0.064	0.050	0.000	0.864
USD	-0.422	0.000	1.678	0.000	0.000	0.000

Note: These are the VAR-LiNGAM model estimates of the causal effect matrices $\mathbf{B}_0$ in the before (A), surrounding (B), after (C), and long after (D) estimation data sets for the Australian dollar (AUD), Canadian dollar (CAD), euro (EUR), Great Britain pound sterling (GBP), Japanese yen (JPY), and the United States dollar (USD). A $\mathbf{B}_0$ matrix contains the causal affects within the lag 0 period.

Table C-6. Estimation Data Set Separating Matrices

A

	AUD	CAD	EUR	GBP	JPY	USD
AUD	1041.500	-975.250	-799.850	192.680	-1058.700	763.400
CAD	835.720	-1312.800	686.570	-517.450	418.740	1065.300
EUR	305.090	264.360	-1718.700	1449.600	403.100	-359.300
GBP	-748.980	-660.030	931.700	1437.400	-462.980	-53.321
JPY	813.100	558.800	260.040	216.830	246.000	-1831.200
USD	-338.380	1108.200	238.330	-428.490	-722.210	796.750

B

	AUD	CAD	EUR	GBP	JPY	USD
AUD	-1056.600	866.120	419.950	-105.300	995.360	-540.200
CAD	-261.170	1219.800	-1543.200	1237.900	-114.950	-1134.300
EUR	-0.382	759.010	-36.187	-1803.300	-143.010	498.620
GBP	-1134.700	663.180	1131.100	1.400	-933.740	258.420
JPY	384.990	464.420	620.990	-178.290	429.420	-1955.600
USD	431.480	927.850	-388.920	-380.770	8.164	146.330

C

	AUD	CAD	EUR	GBP	JPY	USD
AUD	-841.820	1196.200	-857.140	1270.100	667.790	-1341.300
CAD	171.540	-575.170	-23.868	1574.200	-821.910	-186.330
EUR	-1066.100	786.610	1535.100	-637.470	-131.120	-202.190
GBP	-206.140	922.530	-551.850	-173.550	-1187.200	666.690
JPY	-377.760	-364.880	-185.320	58.571	-505.290	1772.500
USD	708.380	741.860	-633.710	-357.350	171.970	-143.690

D

	AUD	CAD	EUR	GBP	JPY	USD
AUD	160.910	44.985	-2068.500	-138.080	-59.670	60.764
CAD	1550.900	-1872.500	-104.360	-232.590	-224.620	945.420
EUR	163.100	-60.696	-1500.400	2116.700	-31.328	-489.860
GBP	-699.070	-1096.900	975.310	843.040	380.600	-703.930
JPY	-222.460	-341.220	-157.180	-550.090	600.320	955.210
USD	350.780	246.930	-331.350	-40.946	1642.000	-1799.500

Note: These are the independent component analysis estimates of the separating matrices ***B*** in the before (A), surrounding (B), after (C), and long after (D) estimation data sets. A separating matrix facilitates the computation of the independent components from the original series of returns.

Table C-7. AR Model Parameter Estimates

Currency	Lag(1) Parameter	Constant
Before Estimation Data Set		
AUD	-0.100	0.006
CAD	-0.061	-0.009
EUR	-0.033	-0.009
GBP	-0.014	-0.004
JPY	-0.001	0.006
USD	0.013	-0.009
Surrounding Estimation Data Set		
AUD	-0.090	-0.008
CAD	-0.066	0.010
EUR	-0.049	0.018
GBP	-0.030	-0.004
JPY	-0.007	0.011
USD	0.036	-0.018
After Estimation Data Set		
AUD	-0.074	0.002
CAD	-0.062	-0.011
EUR	-0.047	-0.006
GBP	-0.018	-0.012
JPY	0.007	0.000
USD	0.023	-0.002
Long After Estimation Data Set		
AUD	-0.136	0.001
CAD	-0.098	0.005
EUR	-0.044	0.007
GBP	-0.012	-0.009
JPY	0.003	0.001
USD	0.038	-0.003