Essays on Asset Allocation and Econometrics

Marco Aiolfi

December 20, 2005

Acknowledgements

I would like to thank Carlo Favero and Allan Timmermann, my main advisors, for their suggestions and constant support during my research. Carlo Favero first introduced me to econometrics when I was an undergrad at Bocconi University. He supervised my dissertation and pushed me to go further and enroll in a Ph.D programme and spend time abroad at UC San Diego as a visiting student. Without Carlo Favero I would not be here today. Allan Timmermann accepted to supervise me during my visit at UC San Diego. He shaped my knowledge of finance and econometrics and always pushed me to the limit. I learnt a new way of working from him and for this I'm grateful. The third member of my committee, Bruce Lehmann, expressed interest in my work during my visit to UC San Diego and shared with me his encyclopedic knowledge at Starbucks. Bruce is probably one of the smartest economist I have ever met in my life.

During my visit to UC San Diego I had the pleasure of meeting some of the best econometricians in the world. I am particularly thankful to Graham Elliott and Clive Granger for their time. Special thanks go also to Pierpaolo Battigalli, Luis Catao, Francesco Giavazzi, Ivana Komunjer, Massimiliano Marcellino, Maurizio Motolese, Fulvio Ortu, Luca Sala, Rossen Valkanov.

I should also mention my graduate fellows at Bocconi University and UC San Diego. I want to thank Andrea Carriero, George Monokroussos, Carlos Capistran Carmona, Kevin Sheppard for

Abstract

This thesis develops and implements new econometric approaches to answer interesting questions in finance and macroeconomics.

In my job market paper, *What Drives Money Managers' Portfolios? An Investigation of Preferences and Beliefs*, I propose a general framework to identify and estimate the parameters characterizing the preferences and beliefs of money managers. In a mean-variance framework we provide joint estimates of the preference parameters and the beliefs conditioned on observable information that most closely reproduce the dynamics of the observed portfolio recommendations made by a panel of international money managers for *The Economist*. Our findings suggest that money managers behave as low risk-averse investors and that heterogeneity in their conditional beliefs is key to explaining differences in the recommended portfolio allocations. The source of heterogeneity lies in the diverse interpretation of publicly available information. When we test for rationality, we can reject the hypothesis that investment banks are using information efficiently only in a few cases.

In *Persistence in Forecasting Performance and Conditional Combination Strategies* we consider measures of persistence in the (relative) forecasting performance of linear and nonlinear time-series models applied to a large cross-section of economic variables in the G7 countries. We find strong evidence of persistence among top and bottom forecasting models and relate this to

countless discussion on different topics. Marius del Giudice Rodrigues shared with me his passion for finance and had a special role in these years, being first of all a great friend.

I spent the last three years of my life in San Diego. The life there would have not be the same without Lance and Yenni Larson, BJ and Marion Adamson, Martin and Kristine Bacich, Annalisa Radeghieri, Andrea Assanelli, Andrea Pompa, Damian Bacich, Matt Cooling, Barry and Laura Stohlman, Lorenzo Albacete, Filippo and Fabrizia Marchioni.

Special thanks to Bisi, Francesco, Roberto, Claudio, Maurizio, Don Mauro, and Vitta for their friendship and continuous support.

Of course, I am grateful to my parents for their unconditional support and for accepting that their son was to stay away. Without them and Michela this work would never have come into existence (literally).

Milan, Marco Aiolfi

December 20, 2005

the possibility of improving performance through forecast combinations. We propose a new four-stage conditional model combination method that first sorts models into clusters based on their past performance, then pools forecasts within each cluster, followed by estimation of the optimal forecast combination weights for these clusters and shrinkage towards equal weights. These methods are shown to work well empirically in out-of-sample forecasting experiments.

In *Thick Modelling, Model Uncertainty, and the Predictability of Stock Returns*, we consider the results contained in Pesaran-Timmerman (1995), which provided evidence on predictability of excess returns in the US stock market over the sample 1959-1992. We show that the extension of the sample to the nineties weakens considerably the statistical and economic significance of the predictability of stock returns based on earlier data. We propose an extension of their framework, based on the explicit consideration of model uncertainty under rich parameterizations for the predictive models. We propose a novel methodology to deal with model uncertainty based on "thick" modeling, i.e. on considering a multiplicity of predictive models rather than a single predictive model. We show that portfolio allocations based on a thick modeling strategy systematically outperforms thin modeling.

In *Common Factors in Latin America's Business Cycles* we construct new business cycle indices for Argentina, Brazil, Chile, and Mexico based on common dynamic factors extracted from a comprehensive set of sectoral output, external trade, fiscal and financial variables. The analysis spans the 135 years since the insertion of these economies into the global economy in the 1870s. The constructed indices are used to derive a business cycle chronology for these countries and characterize a set of new stylized facts. In particular, we show that all four countries have historically displayed a striking combination of high business cycle volatility and persistence relative to benchmark countries, and that such volatility has been time-varying, with important differences across policy regimes. We also uncover a sizeable common factor across the four economies which has greatly limited scope for regional risk sharing.

Contents

Chapter 1

What Moves Money Managers' Portfolios? An Investigation of Preferences and Beliefs

1.1 Introduction

For the majority of investors the most fundamental portfolio decision is how to allocate money among equities, bonds, and cash. Nevertheless, almost no attention has been paid to examining professional advice on this issue. This is surprising, since asset allocation is recognized as a major determinant of risk and return and is therefore one of the primary services provided by most brokerage firms and investment advisors. *'Asset allocation is the area in which you can exert the most influence over your returns'*, says David Darst, chief investment strategist for Morgan Stanley's Individual Investor Group.[1] The standard asset allocation literature has focused on the

[1] *Business Week*, 27 June 2005, Investment Guide, p.86.

problem of a representative agent who must allocate his wealth among broad asset classes, quite possibly with different investment horizons, proposing several ways to optimally solve the problem (see Campbell and Viceira (2002) and Brandt (2005) for a review). Little work has been done on the asset allocation choices and the factors influencing market timing. Researchers investigating the performance of institutional investors (Carhart (1997), Blake, Lehmann, and Timmermann (1999)) or published portfolio recommendations (Barber and Loeffler (1993), Canner, Mankiw, and Weil (1997), Chevalier and Ellison (1997), Graham and Harvey (1996)) examine whether these experts possess a superior timing ability, yet they never address the issue of what drives their portfolio recommendations. Investors should understand how money managers implement their investment policies dynamically over time. How, for example, does a bank's equity or bond exposure change in a time of high interest rates or market exuberance? An understanding of these patterns is particularly important for investors who may choose to implement a portion of their portfolio strategy using professional advisors' recommendations.

To identify the drivers of money managers' portfolios, we need to reverse engineer their portfolio recommendations. Our research is linked to the so-called recoverability problem, which is concerned with the possibility of recovering an investor's preferences by observing her consumption/investment decisions. The relevance and apparent difficulty of the task is well summarized by Kraus and Sick (1980) : *'Since individual agent optimality conditions involve the product of probability and marginal utility, it may be that any set of equilibrium prices that are consistent with some combination of beliefs and preferences could also have resulted from different beliefs combined with different preferences'*. Furthermore, in the presence of delegated portfolio management, the portfolio choice that optimally trades off risk and return for investors may differ from the choice that maximizes money managers' utility. It is likely that career concerns and herding behavior play a crucial role in determining optimal portfolio composition. In this respect our work is related to the literature studying optimal design of compensation schemes of fund managers and delegated portfolio management.[2]

To the best of our knowledge, there is no previous research that tries to completely reverse engineer the problem of portfolio choice. Wolf and Pohlman (1983) consider the problem of a bond dealer who has to choose the optimal composition and size of the bid. They attempt to recover his risk aversion by using his actual demand for bills and the distribution of bond returns calculated from the forecasts made by the dealer in weekly auctions. In a mean-variance framework French and Poterba (1991) ask what set of expected returns would explain the observed pattern of international portfolio holdings given an in-sample estimate of the covariance matrix, but they ignore the joint determination of parameters characterizing beliefs and preferences. Dybvig and Polemarchakis (1981), Dybvig and Rogers (1987), Wang (1993), and Cuoco and Zapatero (2000) examine from theoretical standpoint the extent to which a given consumption/investment plan can be rationalized by a simultaneous choice of an agent's preferences and beliefs. Hansen and Singleton (1982) study a representative investor with power utility and develop methods for estimating preferences from the investor's Euler equations. Previous research tried to reduce the complexity of the problem and avoid the identification issues either by focusing solely on one of the two determinants of portfolio weights (preferences or beliefs) or by approaching the recoverability problem only from a theoretical point of view.

In this paper we address whether there is information in the portfolio holdings not contained in the portfolio returns. Furthermore, can we use this information to recover the preferences of money managers and their conditional beliefs about the market? Portfolio weights are the decision variables in the money managers' optimization problem, so it is natural to focus on portfolio weights if we wish to understand the driving forces of investment banks' portfolios. Focusing on portfolio returns ignores potentially useful information that is often available: the composition of the managed portfolio. We address these issues by focusing on the asset allocation strategies provided

[2]See for example Basak, Pavlova, and Shapiro (2004), Goetzmann, Ingersoll, and Ross (2003), Huddart (1999).

by a panel of international investment houses between 1981 and 2005 and regularly published in the *Financial Report*, a confidential newsletter purchased by *The Economist* in 1989. In these surveys money managers are asked to provide asset allocations among equity, bond and cash for an hypothetical investor *'with no existing investments, no overriding currency considerations and an objective of long term capital growth'*. The data set provides us with a unique opportunity to study what drives the recommended portfolios at the asset allocation level and thus allows us to characterize and quantify the investment strategy of a key group of money managers. We develop an integrated approach in order to simultaneously estimate the money managers' preferences and the parameters driving their beliefs on the future state of the market. We show that if the money managers' behavior is optimal (i.e. they use the first order conditions of the optimization problem to determine the portfolio weights), the recoverability problem is well defined. Given some regularity conditions on the utility function and on the functional form of the portfolio policy functions we establish conditions under which we are able to back out the money managers' preferences and their conditional beliefs. A major difference in our approach compared to that of the related literature is that we treat the portfolio policy rules as the outcome of a possible misspecified model, building on the econometric framework developed by Elliott, Komunjer, and Timmermann (2005). Practical implementation of our approach requires the inversion of portfolio weights as well as knowledge of their functional form. Our empirical application is based on the classical mean variance framework that enables us to derive analytical results and closed form solutions. Nevertheless, the integrated approach proposed here is much general and can be applied to other utility specifications.

Using the portfolios recommended in *The Economist* Portfolio Poll, our approach provides very sensible estimates of the money managers risk aversion and sheds light on the state variables used by the different investment houses to form their expectations on the evolution of markets. The point estimates suggest that money managers behave as low risk-averse investors with an average risk aversion coefficient of two, while Swiss banks display an above average risk aversion. Looking at the cross-section of money managers, we find that macro (inflation, growth in industrial production) and financial (short term risk-free rate, default spread) variables are included in the majority of their portfolio policy functions. According to the literature on performance evaluation, we also find evidence of a momentum effect in the recommended investment strategies. The relative performance of the stock and bond market in the previous period is found to influence money managers' beliefs on the future state of the markets. Furthermore, we find that heterogeneity in the beliefs is key to explaining differences in the recommended portfolio allocations. Indeed, the center of their disagreement is the diverse interpretations of the public available information. For example, we find disagreement among the money managers regarding the asset allocation implications of high level of inflation. For some of the investment houses participating in the survey, none of the macro or financial variables used to track time-varying risk premia is found to be significant. Finally, we ask if money managers are rational when they construct the recommended portfolios, i.e if they are using public information efficiently. Overall we can reject rationality in 10%-25% of the cases depending on the variable entering the portfolio policy function and the instruments' set.

The remainder of the paper is organized as follows. In the next section we introduce the asset allocation problem and some notations. Section 3 outlines the econometric approach adopted in the paper and provides details on the identification conditions and the estimation strategy. Section 4 describes in detail the money manager's problem in a mean variance framework and sets up the basis for the estimation. Section 5 discusses the nature and content of the *The Economist Portfolio Poll*. Section 6 summarizes the empirical findings and Section 7 concludes the paper.

1.2 The Asset Allocation Problem

In this section we study the money managers' objectives and establish conditions under which we can jointly identify the parameters describing their utility function and conditional beliefs on future investment opportunities from a sequence of observed portfolio recommendations. As expected, identification turns out to be a key issue if we want to extract information from the portfolio weights.

1.2.1 The Setup

Consider the problem of a money manager who has to recommend a portfolio strategy for a buy and hold investor who is allocating his wealth W_t among m financial risky assets and one risk-free asset. Denote by $R_{i,t+h}$ the random gross excess return on the i-th risky asset for $i = 1, ..., m$ and by $R_{t+h} = (R_{1,t+h}, ..., R_{m,t+h})'$ the $m \times 1$ vector of excess returns for period $t + h$. Let $\Omega_t^{(j)}$ be the information set of the j-th money manager, with $j = 1, ..., J$. Let $X_t \in \Omega_t^{(j)}$ denote the $k \times 1$ vector of observed state variables that the j-th money manager considers when making his asset allocation decision. Denote by $F_{R|X}$ the conditional distribution of returns R_{t+h} given $X_t = x_t$. Let $\omega_t^{(j)} = \left(\omega_{1,t}^{(j)}, ..., \omega_{m,t}^{(j)}\right)'$ denote the $m \times 1$ vector of portfolio weights recommended by the j-th money manager for period t. We assume that short sales are not allowed and the portfolio weights must sum to one, so the set of admissible portfolio weights is given by $A \equiv \left\{\omega_t^{(j)} \in \mathbb{R}^m : \sum_{i=1}^m \omega_{i,t}^{(j)} = 1, \quad \omega_{i,t}^{(j)} \geq 0\right\}$.[3]

Money managers' preferences are assumed to admit an expected utility representation with a concave utility function $U\left(W_{t+h}, \gamma^{(j)}\right)$ defined over terminal wealth $W_{t+h} = W\left(\omega_t^{(j)}, R_{t+h}\right)$. The vector $\gamma^{(j)}$ captures the parameters characterizing the utility function of the j-th money manager (for example the risk aversion). The money manager uses a model $F_{R|X}(\phi^{(j)})$ of the conditional distribution of the returns given the state variables, where $\phi^{(j)}$ is the vector of parameters driving the evolution in the beliefs about the returns distribution. We collect the model parameters for the j-th money manager in the vector $\psi^{(j)} = \left(\gamma^{(j)\prime}, \phi^{(j)\prime}\right)'$ where $\psi^{(j)} \in \mathbb{R}^p$ is known to the money

[3]This assumption is reasonable since the investment advisors are concerned with the allocation across a stock portfolio, a bond portfolio, and a money market portfolio. For Merrill Lynch, for example, the portfolios can be mutual funds or bond or stock accounts managed directly by Merrill Lynch.

manager but unknown to the researcher.

We can formalize the solution to the asset allocation problem by introducing the definition of a portfolio policy function.

Definition 1.2.1 ***Portfolio Policy Function (or Investment Strategy).*** *A portfolio policy is defined as a function $\omega_t \equiv \omega(x_t, \psi) : \mathcal{X} \times \mathcal{G} \rightarrow \mathbf{A}$ where $\mathcal{X} \subseteq \mathbb{R}^k$ is the range of values of X_t, $\mathcal{G} \subseteq \mathbb{R}^p$ is the range of values of ψ, and $A \subseteq \mathbb{R}^m$ is the set of admissible portfolio weights.*

One possibility for estimating the optimal portfolio strategy ω_t^* is to model it parametrically and seek to solve for the parameters that maximize the expected utility of the money manager. The idea of directly modeling the portfolio weights as functions of observable economic quantities was introduced by Brandt (1999) and then further developed by Ait-Sahalia and Brandt (2001), Brandt and Santa-Clara (2004), Brandt, Santa-Clara, and Valkanov (2005). This strategy, also known as *single step procedure*, is particularly appealing since it is fully integrated into a utility maximization framework and does not require the complete specification of the conditional distribution of the returns. Considering that (1) the optimal portfolio weights are the ultimate object of interest; (2) there is vast disagreement even among financial economists on how to model best returns; (3) the same predictor variables might affect different moments of the returns, it seems natural to focus on the single step procedure. A parametric model for the optimal investment strategy is defined as follows.

Definition 1.2.2 ***Parametric Portfolio Policy Function.*** *A parametric model of the portfolio policy ω_t is a collection of portfolio policies $\{\omega(x_t, \theta(\psi)), \theta(\psi) \in \Theta\}$ where Θ is a compact set. Examples of a parametric model are the constant allocation strategy ($\omega_t = \overline{\omega}$), the linear allocation policy ($\omega_t = \theta(\psi)x_t$), and the non linear allocation policy ($\omega_t = \omega(x_t, \theta(\psi))$).*

In the case of a portfolio of m risky assets, ω_t is the $m \times 1$ vector of investment strategies, *and $\theta(\psi) = [\theta_1(\psi)', \ldots, \theta_m(\psi)']$ is $m \times k$ matrix of coefficients where $\theta_i(\psi)$ for $i = 1, \ldots, m$ is the $k \times 1$ vector of parameters mapping the state variables into the portfolio policy for the i-th risky asset.*

Given a parametric model for the investment strategy, suppose there exists a parameter value, $\theta(\psi) \in \Theta$, that maximizes the expected utility of the money manager within the parametric family $\omega(x_t, \theta(\psi))$. The optimal problem of the j-th money manager can be reformulated as:

$$\begin{aligned} \theta^*(\psi^{(j)}) &= \arg\max_{\theta} E_t \left[U \left(W \left(\omega \left(x_t, \theta(\psi^{(j)}) \right), R_{t+h} \right), \gamma^{(j)} \right) \right] \\ &= \arg\max_{\theta} \int U \left(W \left(\omega \left(x_t, \theta(\psi^{(j)}) \right), R_{t+h} \right), \gamma^{(j)} \right) dF \left(R_{t+h} | x_t; \phi^{(j)} \right). \end{aligned} \tag{1.1}$$

The optimal portfolio allocation, $\omega_t^{(j)^*} = \omega(x_t, \theta^*(\psi^{(j)}))$, suggested by the j-th money manager depends upon X_t through the subjective conditional distribution of returns $F_{R|X}\left(\phi^{(j)}\right)$, and upon $\gamma^{(j)}$, and could differ across money managers due to:

1. differences in the conditioning state variables, $X_t \in \Omega_t^{(j)}$ (information disparity);
2. differences in the conditional subjective probability distribution of the returns $F_{R|X}\left(\phi^{(j)}\right)$ (belief disparity);
3. differences in the parameters $\gamma^{(j)}$ characterizing the utility function (preference disparity).

The superscript j is henceforth suppressed and we will simply refer to the generic money manager.

1.3 The Econometric Framework

Our integrated approach to reverse engineer the portfolio weights differs from those adopted by earlier studies on the recoverability problem in several ways. First, we focus on the portfolio weights, and not on the portfolio's return. Besides the obvious fact that the optimal portfolio weights are the ultimate object of interest, there is another benefit from focusing on the portfolio weights. We are able to exploit the information on the composition of the managed portfolio which is often ignored by the literature on the recoverability problem. Second, the approach developed here is much more general since it allows the joint identification and estimation of both the preferences and conditional beliefs implied in an observable sequence of portfolio recommendations or holdings. Previous studies tried to reduce the complexity of the problem and avoid the identification issues by focusing only on one of the two determinants of portfolio weights (preferences or beliefs).[4] Wolf and Pohlman (1983) consider the problem of a bond dealer who has to choose the optimal composition and size of a bid and try to recover his risk aversion by using his actual demand for bills and the distribution of bond returns calculated from the forecasts made by the dealer himself for the weekly auction. In a mean-variance framework French and Poterba (1991) ask what set of expected returns would explain the observed pattern of international portfolio holdings given an in-sample estimate of the covariance matrix and a particular coefficient of risk aversion. Hansen and Singleton (1982) consider the idea of backing out the parameter values that are consistent with an optimizing agent's objective function. They develop a method for estimating preference parameters from the investor's Euler equations. A third difference between previous research and our approach is that they treat asset returns or portfolio weights as observable state variables. In our case we treat the portfolio weights as the outcome of a possibly misspecified econometric model, i.e. the parametric

[4] Lehmann (2005) asks what we can learn about beliefs from probability statements about sample moment conditions in rational expectations models, under the maintained hypothesis that the moment conditions are true and the actions of the decision maker cannot affect the random evolution of the state vector.

portfolio policy function.[5] Building on the econometric framework developed by Elliott, Komunjer, and Timmermann (2005) we then establish conditions on the money managers' portfolio policy function under which the parameters describing preferences and beliefs are identified and can be estimated.

1.3.1 Optimal Portfolio Weights

The relevant optimality condition for the money managers' decision problem in (1.1) is given in the following Proposition. Assumptions referred to in the propositions are listed in Appendix A while proofs are provided in Appendix B.

Proposition 1 (Optimality) *Under assumptions [A1]-[A6] and for given $\psi_0 \in \mathbb{R}^p$ the policy function is optimal if and only if*

$$E_t \left[\nabla_\theta U \left(W_{t+h} \left(\omega^* \left(X_t, \theta^*(\psi) \right), R_{t+h} \right) \right) \right] = 0 \tag{1.2}$$

where $\nabla_\theta U(\theta) = [J_\theta \omega_t \nabla_\omega W_{t+h} \frac{\partial U(W_{t+h},\psi)}{\partial W_{t+h}}]$ and $J_\theta \omega_t$ is the jacobian of the portfolio weights with respect to θ. Moreover, given ψ the solution ω_t^* to the $(m \times 1)$ system of orthogonality conditions in (1.2) is unique, and the implicit function $\omega_t^* = \omega\left(X_t, \theta^*\left(\psi\right)\right)$ is a continuously differentiable one-to-one mapping from $G \subseteq \mathbb{R}^p$ to $A \subseteq \mathbb{R}^m$.

Proposition 1 shows that under fairly weak assumptions on θ, $U(\cdot,\cdot)$, and the joint distribution of R_{t+h} and X_t, the sequence of optimal investment strategies, ω_t^* satisfies the moment conditions in (1.2). Given ψ, if the money manager uses (1.2) to determine the optimal investment strategy, ω_t^*, then for a given ω_t^* we can back out ψ by using the same condition. However this approach is valid only if knowing a solution to (1.2) allows the econometrician to identify ψ. When the

[5] The misspecification we have in mind concerns the functional form and the variables included in the portfolio policy function. Further details are given below.

portfolio policy is optimal any information must be correctly included in ω_t^* and the quantity inside square brackets in (1.2) is a martingale difference sequence. The second part of Proposition 1 shows the existence of a unique solution $\omega(X_t, \theta^*(\psi))$ to the asset allocation problem that in turn, knowing ω_t^* yields a unique value of ψ. Without this relationship we would not be able to identify ψ. Proposition 1 is very general and allows for non-linear policy rules, $\omega_t \equiv \omega(X_t, \theta(\psi))$, provided that the policy rule is identifiable for each realization of the forecasting variable X_t (i.e. [A5] holds). If in addition $\omega(X_t, \theta(\psi))$ is twice continuously differentiable and convex in the parameter θ, then ω_t^* is an optimal investment strategy.

1.3.2 Identification

The researcher observing the recommended investment strategies does not know exactly which state variables enter the information set of the money manager. In particular we would expect that money managers have access not only to publicly available information but also to private information which is not available to the researcher. For example, money managers may have private information coming from their web of relationships with companies' managers and financial analysts. To get identification we need to impose some restrictions on the functional form of the portfolio policy function. We focus on $J_\theta \omega_t$, the jacobian of the portfolio policy functions with respect to the parameter vector, θ. In the linear case the generic element of the jacobian is simply X_t, but for non-linear policy rules it could potentially depend on both X_t and the entire vector of parameter values θ^*, which is not available to the econometrician observing the recommended portfolio strategies. In order to get identification we need to know the functional form of the portfolio policy function, the true values of the parameters as well as all the values of the variables in the information set of the money managers.

We now consider some examples to clarify the restrictions on the portfolio policy function necessary to get identification. Suppose that X_t, the vector of state variables considered by the money manager when making his asset allocation recommendations is given by $X_t = \{Y_t, Z_t\}$ where Y_t is private information available only to the money manager and Z_t is public information.

Example 1 *If the optimal portfolio policy function is a linear function of the elements in X_t or it is separable in Z_t, then the jacobian is exactly X_t and knowledge of Z_t is sufficient to guarantee identification. The portfolio policy $\omega_t^* = \theta_1^*(\psi)Y_t + \theta_2^*(\psi)Z_t$ is admissible since knowledge of public information, Z_t, is sufficient to guarantee identification of ψ.*

Example 2 *If the optimal portfolio policy has the form $\omega_t^* = \theta_{12}^*(\psi)Y_tZ_t$, then the jacobian is given by Y_tZ_t and knowledge of both Y_t, and Z_t is necessary to achieve identification. Since Y_t is private information available only to the money manager, we are not able to get identification of the vector of parameters of interest.*

In general, under separability of the portfolio policy function in the parameters and in the state variables, moment conditions based on a sub-vector $Z_t \in X_t$ are sufficient to identify ψ. Given this practical concern, in the remainder of the paper we will focus on linear portfolio policy functions $\omega_t = \theta(\psi)x_t$. The assumption that the optimal portfolio weights are linear functions of the state variables is innocuous because one can think of the linear policy function as a more general portfolio policy that can be spanned by a polynomial expansion in a more basic set of state variables.[6] Indeed X_t can include non-linear transformations of a set of more basic state variables $\widetilde{X}_t$. In other words, our approach can in principle accommodate very general dependence of the optimal portfolio weights on the state variables.

The next proposition formalizes the intuition of the previous examples and shows that moment conditions based on an observed subvector Z_t of X_t are sufficient to identify ψ, provided

[6] See Brandt (2005), Brandt and Santa-Clara (2004), Brandt, Santa-Clara, and Valkanov (2005) for a detailed discussion on this point.

that the investment strategy and the wealth dynamics are linear.

Proposition 2 *Under assumptions [A1]-[A7], and linearity of the portfolio policy function, given a solution $\omega_t^* = \theta^*(\psi)X_t$ to (1.2), the true value ψ is the unique minimum of a quadratic form*

$$Q_0(\psi) = E_t\left[Z_t R_{t+h} \frac{\partial U(W_{t+h}, \psi)}{\partial W_{t+h}}\right]' S^{-1} E_t\left[Z_t R_{t+h} \frac{\partial U(W_{t+h}, \psi)}{\partial W_{t+h}}\right] \tag{1.3}$$

where Z_t is a subset of X_t and S is any positive definite weighting matrix.

An important implication of the previous proposition is that in order to back out ψ the econometrician does not need to use the full vector of variables, X_t, used by the money manager. Moment conditions based on a sub vector of these variables, Z_t, are sufficient to identify ψ. Since $Q_0(\psi)$ is a quadratic form, the minimum is unique and using Z_t instead of X_t will affect only the curvature of the function and thus the precision of the estimates. This result is rather strong and grants that only with public information we can identify the parameters driving the preferences and beliefs of the money managers, even though we do not know the full portfolio policy model used to create the investment strategy. In our application the result in Proposition 2 is crucial since it allows for misspecification in terms of functional form and variables included in the investment strategies recommended by the investment banks. The next section deals with some econometric issues arising in our approach.

1.3.3 Implementation

In practice we only observe the sequence of portfolio weights $\{\widehat{\omega}_t\}$ recommended by a money manager, where $\widehat{\omega}_t \equiv \widehat{\theta}_t x_t$ and $\widehat{\theta}_t$ is an estimate of θ^* obtained by using the data up to time t. Let T be the total number of periods available and assume that the first τ observations are used to produce the first set of portfolio weights $\widehat{\omega}_\tau$. There are $T - \tau$ set of portfolio weights available, starting at $t = \tau$, and ending at $t = T$. The portfolio weights are assumed to be constructed recursively so that the parameter estimates use all information prior to the period for which the weights are chosen. In particular the investment strategy $\widehat{\omega}_{\tau+i}$ is constructed using data up to $\tau + i$ to compute an estimate $\widehat{\theta}_{\tau+i}$ of θ^*. This approach allows for the possibility that the money manager is recursively learning the parameters of the portfolio policy function.

Having observed the sequence of portfolio weights $\{\widehat{\omega}_t\}_{\tau \leq t < T}$, we can now construct an estimate of ψ. In practice we need to invert the mapping implied by the optimality condition in (1.2). To do so we need to know the functional form of the portfolio policy functions implied by a particular utility function. Unfortunately in discrete time we are able to get closed form solutions only in very specifical cases. Given this consideration we stay inside the classical mean-variance framework. The main motivation is to take advantage of the analytical tractability. Specifically, it allows us to exploit the affine closed form solution for the portfolio policy functions which otherwise could be highly non linear functions of the preferences and beliefs. This feature is particularly attractive in our specification because it allows us to estimate the parameters by minimizing the errors of the portfolio policy.

The next section describes a mean-variance framework with m risky assets, setting the basis for our estimation strategy.

1.4 Mean Variance Framework

In this section we provide analytical expressions for the optimal weights in a portfolio of m risky assets.

Consider a fund manager with CARA utility

$$U[W_t] = -\frac{1}{\gamma}\exp[-\gamma W_t], \tag{1.4}$$

where γ is the risk aversion and assume that the manager has to choose the optimal allocation among m risky assets and a risk-free asset. The return on the portfolio recommended by the manager is given by

$$\begin{aligned} R^P_{t+1} &= \omega_{1,t}(R_{1,t+1} - rf) + \omega_{2,t}(R_{2,t+1} - rf) + rf \\ &= rf + \omega_t' R_{t+1} \end{aligned}$$

where $\omega_t = (\omega_{1,t}, ..., \omega_{m,t})'$ is the $(m \times 1)$ vector of portfolio weights and $R_{t+1} = (R_{1,t+1}, ..., R_{m,t+1})'$ is the $(m \times 1)$ vector of gross excess returns of the two risky assets over the risk-free asset. Furthermore suppose that the excess return distribution is given by

$$R_{t+1}|\Omega_t \sim N(\mu_{t+1}, \Sigma_{t+1}) \tag{1.5}$$

where μ_{t+1} is the $m \times 1$ vector of expected excess returns and Σ_{t+1} is the $m \times m$ covariance matrix of the returns on the risky assets.

Money managers are in general remunerated on the basis of their absolute performance and their performance relative to a benchmark.[7] There is an open debate about the impact of relative performance considerations and the way they are incorporated in the portfolio weights. In our case there is not any explicit contract between the money manager and the potential investor. In this setting, one could argue that relative performance affect indirectly the reputation of the

[7]In a survey about the compensation structure of portfolio managers, Farnsworth and Taylor (2004) find that most managers bonuses are impacted by investment performance relative to a benchmark and/or peer group.

investment houses.[8] We decided to model the managers' payoff in terms of absolute returns on the recommended portfolio, $W_{t+1} = R^P_{t+1}$ arguing that the best way for the money managers to maximize their reputation is indeed to maximize the return on the recommended portfolio.

The money manager solves the following problem

$$\max_{\omega_t} E_t \left[-\frac{1}{\gamma} \exp \left[-\gamma W_{t+1} \right] \right] = -\frac{1}{\gamma} \exp \left\{ -\gamma E_t \left[W_{t+1} \right] + \frac{\gamma^2}{2} Var \left[W_{t+1} \right] \right\}. \tag{1.6}$$

Using log-normality the solution to this problem is equivalent to

$$\min_{\omega_t} \log \left[-\gamma E_t \left[U \left(W_{t+1} \right) \right] \right] = -\gamma E_t \left[W_{t+1} \right] + \frac{\gamma^2}{2} Var \left[W_{t+1} \right], \tag{1.7}$$

which can be reformulated in the usual mean-variance framework:

$$\max_{\omega_t} \left[rf + \omega_t' \mu_{t+1} - \frac{\gamma}{2} \omega_t' \Sigma_{t+1} \omega_t \right]. \tag{1.8}$$

The first order conditions are given by:

$$\mu_{t+1} - \gamma \Sigma_{t+1} \omega_t = 0 \tag{1.9}$$

and the optimal weights are given by:

$$\omega_t^{MV} = \frac{1}{\gamma} (\Sigma_{t+1})^{-1} \mu_{t+1}. \tag{1.10}$$

[8] Once a firm is established, it is recognized by its name, which is uniquely associated with its characteristics and past performance.

In the specific case with two risky assets we can specialize the solution as:

$$\begin{aligned} \omega_{1,t}^{MV} &= \frac{1}{\gamma} s_{1,t+1} \\ \omega_{2,t}^{MV} &= \frac{1}{\gamma} s_{2,t+1} \\ \omega_{3,t}^{MV} &= 1 - \frac{1}{\gamma} \left(s_{1,t+1} + s_{2,t+1} \right) \end{aligned} \tag{1.11}$$

where

$$\begin{aligned} s_{1,t+1} &= \left[\frac{\sigma_{2,t+1}^2 \mu_{1,t+1} - \mu_{2,t+1} \sigma_{12,t+1}}{\sigma_{1,t+1}^2 \sigma_{2,t+1}^2 - \sigma_{12,t+1}^2} \right], \\ s_{2,t+1} &= \left[\frac{\sigma_{1,t+1}^2 \mu_{2,t+1} - \mu_{1,t+1} \sigma_{21,t+1}}{\sigma_{1,t+1}^2 \sigma_{2,t+1}^2 - \sigma_{12,t+1}^2} \right]. \end{aligned} \tag{1.12}$$

$s_{1,t+1}$ and $s_{2,t+1}$ are conditional information ratios or *beliefs* describing the expected evolution of the investment opportunity set for the two risky assets. To gain economic intuition we can look at $s_{1,t+1}$ ($s_{2,t+1}$) as an indicator of the relative attractiveness of stocks versus bonds (bonds versus stocks). These beliefs are not observable, so we need to build a proxy for the evolution of the investment opportunity set.

Suppose that the money managers use an affine specification for the beliefs

$$s_{i,t+1} = \lambda_{i1} + \lambda_{i2} z_t + \varepsilon_{i,t}, \qquad i = 1, 2 \tag{1.13}$$

where z_t is public available information at time t, when the recommendations are made. This specification for the dynamics of the beliefs guarantee that the policy function is linear in the state variables. Therefore, we can identify the parameters of interest given only a subset, z_t, of the x_t variables entering the information set of the money manager. Furthermore, we can look at the linear representation in (1.4) as a first order Taylor expansion of a more general non-linear process. The interpretation of λ_{12} (λ_{22}) is straightforward. An increase of one unit in z_t gives us a measure of how much more (less) attractive asset 1 becomes with respect to asset 2. In general we would expect λ_{12} and λ_{22} to have the same magnitude and opposite sign: if stocks become more attractive than bonds it is natural that bonds are less attractive than stocks. In practice since covariance terms enter the expression this is not necessarily true. The suggested timing here is as follows: (1) the money managers form their beliefs on the evolution of the stock and bond market and choose the optimal mix accordingly; (2) given their risk aversion they choose the optimal asset allocation between the risk-free asset and the portfolio of risky assets. The composition of the portfolio of risky assets does not help us in understanding the preferences of the money manager. The only way of learning about the risk appetite is to consider the ratio of the proportion invested in the risky assets and the proportion invested in the risk-free rate ($\frac{\omega_{3,t}^{MV}}{\omega_{1,t}^{MV}+\omega_{2,t}^{MV}}$). The classical separation theorem applies: highly risk averse investors should hold more of their portfolio in the risk-free asset, but the composition of risky assets should be the same for all the investors. The two-fund theorem in principle still allows for a good deal of customized portfolio formation if managers have different information or beliefs. Indeed, the knowledge of the composition of the portfolio of risky assets gives us information about the money managers' beliefs summarized by $s_{1,t+1}$, and $s_{2,t+1}$ and their evolution through time.

1.4.1 Estimation

In the mean variance framework outlined above it is possible to invert the optimal portfolio weights by exploiting their closed form solution and proceed to the joint estimation of preferences and beliefs. To do so we treat the recommended portfolio weights as the outcome of a possibly misspecified econometric model by assuming a linear portfolio policy function and minimize the portfolio policy errors. The accuracy of a candidate set of preferences and beliefs can be judged by how well it reproduces the observed portfolio weights. To this end, let us denote by $e_t(\psi)$ the m dimensional vector of portfolio policy errors where $\psi = \{\gamma, \lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22}\}$ is the $p \times 1$ vector of parameters of interest. The idea of our approach is to select parameters that make the sample averages of the portfolio policy errors as close to zero as possible so that

$$\widehat{\psi} = \left\{ \psi : \frac{1}{T}\sum_{t=1}^{T} e_t(\psi) = 0 \right\}. \tag{1.14}$$

Here the vector of portfolio policy errors is given by the difference between the observed portfolio weights and the weights coming from the mean variance optimization problem denoted by $\omega_{1,t}^{MV}$ and $\omega_{2,t}^{MV}$:

$$e_t(\psi) = \begin{bmatrix} \widehat{\omega}_{1,t} - \omega_{1,t}^{MV}(z_t,\psi) \\ \widehat{\omega}_{2,t} - \omega_{2,t}^{MV}(z_t,\psi) \end{bmatrix} = \begin{bmatrix} \widehat{\omega}_{1,t} - \frac{1}{\gamma}(\lambda_{11} + \lambda_{12}z_t + \varepsilon_{1,t}) \\ \widehat{\omega}_{2,t} - \frac{1}{\gamma}(\lambda_{21} + \lambda_{22}z_t + \varepsilon_{2,t}) \end{bmatrix} \tag{1.15}$$

The Generalized Method of Moments (GMM) developed by Hansen and Singleton (1982) gives the econometric framework for the estimation. The GMM estimator of ψ is given by

$$\widehat{\psi} = arg\min_{\psi}[g_T(\psi)' W_T g_T(\psi)], \tag{1.16}$$

where $g_T(\psi) = \frac{1}{T}\sum_{t=1}^{T} e_t(\psi)$ and W_T is a positive definite weighting matrix which may be a function of the data. To identify the parameters we need at least as many moment conditions as there are parameters. In the specific case we have just two moment conditions implied by the economic theory so we need to include instruments, to achieve identification. If the number of instruments is greater than the number required to get identification, the remaining variables can be used to test if the orthogonality condition holds for them, conditioning on the estimated values of the parameters. In a GMM framework this can be done using Hansen's test (or J-test) of over-identifying restrictions. Let v_t be $h \times 1$ vector of instruments. Then the sample counterpart of the orthogonality conditions can be expressed as the $m \times h$ vector $g_T(\psi) = \frac{1}{T}\sum_{t=1}^{T} e_t(\psi) \otimes v_t$. Again the GMM estimator is the value that minimizes the scalar $Q_T = [g_T(\psi)' W_T g_T(\psi)]$. Hansen's J-test is TQ_T and it converges to a chi-squared distribution with $(m \times h) - p$ degrees of freedom. The interpretation is straightforward: the J-test tells us if the money managers are using efficiently all the information contained in the instruments when providing their portfolio recommendations. We will discuss in detail our choice for the instruments in Section 6 when we present the empirical results. For all the estimates presented in the paper the inverse of the estimate of the asymptotic variance of the sample mean of $e_t \otimes v_t$ is used as the weighting matrix[9].

1.5 Data Description

The recommended portfolio weights come from surveys published every six weeks beginning in 1981 in the *Financial Report*, a confidential newsletter purchased by *The Economist* in 1989.[10] *The Economist* continued to conduct the survey but, starting on March 25, 1989, published it every 12 weeks. The nature of the poll is as follows:

> Taking over the portfolio poll that run regularly in Financial Report for eight years, The Economist asked nine money managers for their opinion on the best mix of investments over the next 12 months. They were asked to design a portfolio for an investor with no existing investments, no overriding currency considerations and an objective of long term capital growth. (The Economist, 3/25/1989)

Although the published advice does not necessarily relate to actual managed portfolios, we assume that given the widely respected and read publication outlet, the investment bankers do not take the portfolio poll light-heartedly.[11]

[9]Since we are working with possibly overlapping quarterly data the Newey-West kernel lag length is selected as the $min\left[T^{-\frac{1}{3}}, 3\right]$.

[10]This dataset is ideal for the purpose of the paper since there are not transaction costs which could induce frictions in the dynamics of portfolio weights.

[11]It is well understood that the survey is a worldwide audience. Indeed in the cover letter of the survey is written *'Our poll reaches a readership of over 800,000 subscribers'*.

Each poll comprises three parts. In the first part of the poll the investment houses report their recommended portfolio allocations among equity, bonds and cash.[12] In the second part each institution provides a suggested equity portfolio diversification among North America (United States and others), Europe (France, Germany[13], the United Kingdom (U.K.) and others), and the Far East (Japan and others). Finally in the third part each money manager constructs portfolio of sovereign bonds denominated in six different currencies: US Dollar, Deutsche Mark, Franc French, Yen, Pound Sterling and other. The first part of the poll (asset allocation) was not published between Q3 1997 and Q3 1998, while the third part was first published in August 1997. In this paper we focus on the first part of the poll, the one regarding asset allocation among broad asset classes.

In addition to recommended portfolio holdings, *The Economist* also provides neutral benchmark weights every quarter. These benchmark weights are relative market capitalizations and are only given at the level of security markets. For the neutral asset allocation we use the *Robot Blend* of Arshanapalli, Coggin, and Nelson (2001) which distributes assets 60% to equity, 30% to bonds, and 10% to cash.

Twenty four Houses participated in the surveys from 1981 to 2005.[14] The initial set of recommendations was made by these eight Houses: Anonymous One and Anonymous Two, London Merchant Bank; Brown Brothers Harriman, Wall Street private bankers; Capital House Asset Management; Daiwa Europe, a Japanese investment bank; Scudder Stevens Clark, New York investment counsellors; Phillips and Drew, London brokers; US Trust. Bank Julius Baer, a Swiss

[12]In some instances, the surveys also provided recommended allocations to other assets such as real estate, gold, and fine art. The recommended percentage allocation to these other assets averaged less than 1% throughout the sample. Therefore, we simply added this other allocation to the percentage allocation to cash.

[13]After the survey published in April 2001 *The Economist* no longer reported separate allocations for France and Germany. These countries were now included in a new category, the Euro Area.

[14]The North American Houses are Brown Brothers Harriman, Lehman Brothers, Merrill Lynch and Scudder Stevens Clark. The Asian Houses are Daiwa Europe, Nikko Securities. The European Houses are Anonymous One, Anonymous Two, Commerz International, Bank Julius Baer, Capital House, Credit Agricole, Credit Suisse, Robeco Group, and UBS Phillips & Drew.

investment bank, replaced US Trust in 1983. In 1987, a ninth participant, Wardley Investor Services, joined the group. The set of nine Houses who joined the poll remained unchanged through February 1989. Then, the set of Houses changed periodically through the end of the sample, January 2005.[15] According to the staff of The Economist *'the contributors who have left the poll have chosen to leave or left their positions without designating a substitute. As for selecting banks/asset managers to participate, of course name recognition has counted but they have also allowed smaller participants who satisfied their coverage requirements'*.

The survey is currently conducted via e-mail. Research analysts at each House complete the survey on the Wednesday before the Friday publication date of *The Economist*. In the early days of the sample, however, data-gathering and printing technology might not have allowed for such a quick turn-around. In addition, it is not known exactly when each House set its recommendations. After 1989 the poll is published quarterly but not always before the start of the quarter to which the recommendations apply. The publication date of the recommendations varies from two weeks before to three weeks after the start of the applicable quarter. We account for this problem by computing quarterly returns starting from the recommendation date. We compute quarterly gross returns from the perspective of a US investor assuming no rebalancing during the quarter and we disregard management fees and related costs.

Following common practice, this study uses the Morgan Stanley Capital International (MSCI) value-weighted world index for total returns (capital gains plus dividends) in US dollars as a proxy for the world market portfolio. For bond returns, we use the Merrill Lynch Global Government Bond return index in US dollars, obtained from Global Financial Data. Figure 1 provides a plot of the returns on equity and bonds for the relevant sample period. To represent public information in our empirical application, we use a collection of variables that previous international finance

[15]The Houses not included in the sample are Global Asset Management; Deutschebank; US Trust; Capital Management; Citicorp; Indocam; Rabobank International and Standard Life.

studies find to be useful for predicting time-varying risk premia and volatility.[16,17] The variables are: (1) the quarterly growth rate of G-7 inflation, (2) the quarterly growth rate of G-7 industrial production, (3) the level of the one-month Treasury bill yield, (4), the slope of the US Treasury yield curve, measured as the difference between ten-year and three months fixed maturity bond yields from the CRSP Fama-Bliss files (5) the default premium, computed as the yield spread between Moody's Baa and Aaa rated bonds.[18]

We do not confine our analysis to a study of the cross section of recommended portfolio weights. For this purpose we also compute and analyze what we call the consensus broker: for each broad asset class we compute every quarter an average recommended portfolio weight over all reporting investment firms.

1.5.1 Asset Allocation to Equity, Bonds, and Cash

Table 1 reports descriptive statistics regarding the asset allocation recommendations of the investment houses participating in the poll over the sample 1981-2004. Six Houses made at least 60 recommendations, with Daiwa Europe, Bank Julius Baer, and UBS Phillips and Drew making the most with 106, 91, and 86 recommendations, respectively. On average across the sample period the consensus bank recommended that investors allocate 60% to equity, 31% to bonds, and 9% to cash. The lowest average equity recommendation, 37%, is made by Credit Suisse, the highest, 85%, is issued by Scudder Stevens. Table 2 reports the relative composition of the portfolio of the risky assets and the optimal mix of risky portfolio and the risk-free asset. Credit Suisse, Brown Brothers Harriman, and Capital House display on average the highest holdings in the risk-free asset but also the highest standard deviation in the optimal mix of risky portfolio and risk-free asset.

[16]See for example Campbell (1987), Fama and French (1989), Ferson and Harvey (1991), Pesaran and Timmermann (1995).

[17]For the conditioning variables we use monthly data and match the realization with the month in which the recommendations are issued.

[18]The macro variables are obtained from Global Financial Data, while the interest rates series are from CRSP.

Turning to the optimal mix of risky assets Credit Suisse, Robeco Group, Merrill Lynch, Commerz International, and UBS display a strong preference for bonds versus equity.

Figures 2 and 3 illustrate the time-variation in the asset allocation recommendations for each House, while Figure 4 provides a snapshot of the cross sectional distribution in the recommendations, by computing the average, minimum and maximum recommendation across the Houses participating in the poll at each point in time. Two stylized facts emerge from Figures 2-4. Individual asset allocation recommendations are time-varying and display some volatility, matching the fact that investment opportunities are not stationary. There is a high cross-sectional variation in the recommendations together with a stable and consistent average recommended portfolio closely matching the standard 60/30/10 recommendation. Nevertheless, the average allocation shows the presence of regimes (bull and bear periods); banks were on average more bullish than the static 60/30/10 benchmark between 1982 and 1990. Furthermore, the Houses under-invested in equity and over invested in bonds for a significant time after the Crash of October 1987. There are three spikes in the average allocation to cash: in 1984 following the uncertainty regarding the interest rates and the dollar; in late 1987 following to the decrease in equity holdings after the Crash of October 1987; between September 1990 and April 1991 when the Houses *'were wrong-footed by events in the Gulf'* (*The Financial Report*, September 29, 1990).

We assume that the Houses' portfolio recommendations are made independently. However it is possible that different Houses provide similar portfolio recommendations. Similar suggested portfolios are possible if the Houses share common information and process the information in a similar way. It also might be that relative performance considerations or, more in general, reputation concerns matter and are therefore incorporated in the portfolio recommendations generating herding behavior. In finance there is a huge literature both theoretical and empirical regarding the effects of performance-based compensation schemes[19] and more in general herding

[19]See for example Kapur and Timmermann (2004), Admati and Pfeiderer (1997), Cuoco and Kaniel (2000), Brennan

generated by career/reputation concerns[20] on money managers' portfolio holdings. In general looking at the relation between mutual funds' inflows and performance or at the cross sectional dependence of trading by institutional investors, different studies have provided mixed evidence in favor or against herding.

1.5.2 Portfolio Performance

We computed the return of the recommended investment strategies and compared it to two different benchmark strategies: the *Robot Blend* of Arshanapalli, Coggin, and Nelson (2001) which distributes assets 60% to equity, 30% to bonds, and 10% to cash, and the consensus portfolio which provides information about the average asset allocation of the competitors. Table 3 reports the results. Looking at the managers' portfolio we find that on average over the full sample only two of them have a return which is significantly different from the benchmark returns. Bank Julius Baer under-performs the static (consensus) benchmark by 26 (43) basis points per quarter, while Daiwa over-performs the static (consensus) benchmark by 38 (28) basis points. This is not surprising, since Bank Julius Baer recommends on average a 43% allocation to stocks, compared to the 64% recommended by Daiwa. Interestingly, when we look at the investment banks operating before the October 1987 Crash, also Scudder Stevens and UBS provide a positive return in excess to the benchmark, while Daiwa now offers 104 basis points in excess of the static benchmark and 71 with respect to the competitors. Accordingly to the literature on performance evaluation, we have not been able to find investment advice that statistically outperforms passive or active benchmarks despite the wide cross-sectional variation in asset allocation dynamics.

We also employ the Grinblatt and Titman (1993) Portfolio Change Measure (PCM). The intuition behind the PCM is that informed investors can profit from changing expected returns by

(1993).

[20] See for example Chevalier and Ellison (1997), Goetzmann, Ingersoll, and Ross (2003), Graham and Harvey (1996), Graham (1999), Jaffe and Mahoney (1999).

increasing (decreasing) their holdings of assets whose expected returns have increased (decreased). The holding of an asset that increases with an increase in its conditional expected rate of return will exhibit a positive unconditional covariance with the asset's returns. The PCM is defined as $\frac{1}{T}\sum_t \sum_i [R_{i,t}(\omega_{i,t} - \omega_{i,t-1})]$ where $\omega_{i,t}$ is the portfolio holding of asset i at the beginning of period t and $R_{i,t}$ is asset is return over period t. Yet, if the money manager has appropriately identified which assets will achieve higher or lower returns, the recommended portfolio will exhibit positive covariance between asset returns and portfolio changes. This implies a positive PCM, which serves as evidence of market timing. The Lagged Momentum Measure (LMM) is closely related to the PCM, differing only with regard to the timing of returns. The LMM estimates the covariance between weight changes and previous asset returns and it is defined as $\frac{1}{T}\sum_t \sum_i [R_{i,t-1}(\omega_{i,t} - \omega_{i,t-1})]$. This performance statistic is designed to measure momentum investing by portfolio advisers. The LMM evaluates whether investors shift their recommended portfolio compositions in favor of assets that have recently experienced high returns and away from assets that have underperformed. In Table 3 we compute PCM and LMM for the total portfolio and for the investment in each asset class (equity or bond) recommended by each investment house. It is interesting to notice that only two investment banks display a negative PCM, although of the remaining twelve, only Bank Julius Baer, Capital House and Nikko have a positive PCM at the 10% level. Turning to the LMM numbers, only Credit Agricole and Capital House have positive and significant values. Considering separately the investment in equity and bond does not change the picture. Consistent with the mutual fund literature and previous studies using the same data set (Annaert, Ceuster, and Hyfte (2004), Bange, Khang, and Miller (2004)) we find little evidence of market timing skills at the asset allocation level. These findings make even more interesting the question asked by this paper: what moves money managers' portfolios?

1.6 Empirical Results

In this section we discuss the results obtained with our methodology using the asset allocation recommendations from *The Economist Portfolio Poll*. Table 4 reports the results of iterated GMM estimates of the risk aversion of the money managers, while Tables 5-6 report the estimates of the parameters driving their conditional beliefs about the market.[21] We normalize $\lambda_{11} = 1$ and demean and standardize all the state variables to ease the interpretation of the coefficients of the portfolio policy function. The moment conditions used for the estimation are

$$E\left[\begin{pmatrix} \widehat{\omega}_{1,t} - \frac{1}{\gamma}\left(1 + \lambda_{12} z_t\right) \\ \widehat{\omega}_{2,t} - \frac{1}{\gamma}\left(\lambda_{21} + \lambda_{22} z_t\right) \end{pmatrix} \otimes \begin{pmatrix} 1 \\ z_{t-1} \end{pmatrix}\right]. \tag{1.17}$$

In this case we have two equations and four parameters so we need two instruments to achieve exact identification. Our instruments set comprises a constant and the lagged value of the state variable entering the portfolio policy function, z_{t-1}. We report estimates of the parameters of the portfolio policy function based on different state variables that previous finance literature has found to be a good proxy for time varying investment opportunity set: G-7 Inflation, growth in G-7 Industrial Production, stochastically de-trended risk-free rate, default premium, term spread and the return differential between equity and bonds.[22]

At a first glance a very interesting result emerges: the estimates of the risk aversion parameter are always significant and sensible with values ranging from 1.12 to 2.74 across different banks and state variables.[23] Becker, Ferson, Myers, and Schill (1999) use mutual funds' returns from the CRSP database to simultaneously estimate the risk aversion of a fund manager with CARA utility and the precision of the fund's market-timing signal. For the asset allocation category of funds

[21]The use of multiple starting values provided results nearly identical to the ones reported.

[22]See for example Campbell (1987), Fama and French (1989), Ferson and Harvey (1991), Pesaran and Timmermann (1995).

[23]For a summary of the risk aversion estimates found in previous studies see Bliss and Panigirtzoglou (2004).

the mean value of the fund specific risk aversion is 93.6 with a negative median, -13.4. If so more than half the fund managers had negative risk aversion resulting in a strictly convex utility. Using the same data, and still focusing on the portfolio returns and not on the portfolio holdings, Foster and Stutzer (2003) find more reasonable estimates with a mean value of about 8 compared to our mean value of about 2. Our estimates are robust (in terms of ranking and magnitude) to the inclusion of different state variables in the portfolio policy. Independently of the conditioning variables, Scudder Stevens, Brown Brothers Harriman, and Capital House are the least risk averse while Credit Suisse, Bank Julius Baer, and UBS Phillips are the more risk averse among the Houses in our sample. Interestingly the investment Houses displaying an above average risk aversion are Swiss banks, *'perhaps partly reflecting traditional Swiss preference for fixed interest paper'*, as noted in *Financial Report*, March 17, 1983.

Turning to the estimates of the parameters driving $s_{1,t+1}$ and $s_{2,t+1}$, the beliefs concerning the relative attractiveness of equity and bonds, several interesting results emerge. Having normalized λ_{11} to 1, the value of λ_{21} becomes informative on the money managers' prior on the relative attractiveness of bonds versus stocks. Values of λ_{21} between zero and one suggest that money managers perceive bonds as a less attractive investment compared to stocks. Interestingly the more risk averse banks display values of λ_{21} close to one and in some cases (Bank Julius Baer and Credit Suisse) greater than one. Daiwa Europe and Bank Julius Baer are the only banks that over the full sample significantly outperform and undeperform the competitors. Looking at their estimated values of λ_{21} we notice that Bank Julius Baer have a coefficient twice as big as Daiwa and greater than one. To answer the question of what moves money managers' portfolios, we need to look at how different Houses implement their investment policies dynamically over time. In particular, in the presence of return predictability, the values of λ_{12} and λ_{22} tell us whether the money managers engage in market timing strategies based on the state of the economy and how they incorporate it into the portfolio policy functions. Table 5 summarizes the main findings reporting (whether significant) the sign with which different state variables affect the investment strategies of the money managers. We measure the state of the economy according to macro and financial variables. Inflation and growth in industrial production affect the portfolio policies of different investment houses like Brown Brothers, Scudder Stevens, Capital House, and Daiwa Europe. This finding supports the conclusion of Lamont (2001) who observes that a portfolio tracking the growth rate of industrial production earns positive abnormal returns. Furthermore, industrial production growth enters with the expected sign: an increase in industrial production growth makes equity more attractive than bonds. High inflation is usually bad news for stocks and long-term bonds. However since our specification for the beliefs is in terms of relative attractiveness, we can infer from the estimates that, conditional on high inflation, bonds are perceived as a better investment than stocks for the majority of banks with the exception of UBS. Among the variables tracking the state of financial markets the stochastically detrended short term interest rate and the default spread seem to be particularly important, while the term spread is not incorporated in the portfolio policies in our sample. High levels of short term interest rates usually imply low stock returns and high and volatile short term bond returns with a gradual transition as we move from shorter maturity to riskier and longer term bonds. This mixed effect of the short term rate is captured by our estimates displaying a degree of heterogeneity among the Houses' responses. It seems that the Houses active before the nineties increase their exposition to stocks and lower the positions in bonds in presence of high interest rates, while the opposite is true for Lehman Brothers or UBS. High credit spreads make stocks more attractive than bonds with again Lehman Brothers and UBS being the only managers going against the consensus. The return differential between equity and bonds turns out to be significant and with the expected sign for Brown Harriman, Capital House, and Daiwa. This evidence seems to support the hypothesis of *momentum* strategies in which the Houses move their portfolio based on past return performance of the constituent assets.[24] Finally

[24] Bange, Khang, and Miller (2004) using the same data find evidence of momentum trading for asset allocation to equity and cash.

we notice that none of the listed state variables seem to be incorporated, if marginally, in the policy functions of Commerz International, Credit Suisse, Credit Agricole, and Robeco Group which are also among the most risk averse investment houses. Interestingly those banks participated in the poll only after the nineties. This evidence supports the findings of the predictability literature which documented instability in the relation between returns and forecasting variables.[25] In general it seems that heterogeneity in the conditional beliefs is more important than heterogeneity in the risk aversion to explain differences in the recommended portfolio allocations. Moreover, the source of heterogeneity seems to be in the diverse interpretation of the available information. To support this finding, Figure 6 plots the average fitted values of $s_{1,t+1}$ computed across policy functions based on different state variables for selected investment banks. We choose Daiwa Europe and Bank Julius Baer because they are the only banks that on the full sample significantly outperform and undeperform the competitors and the static 60/30/10 benchmark. We also included UBS because it spans the same sample as the previous banks. As outlined above, UBS usually provides a different interpretation of the signals on the state of economy compared with the other participants in the poll. Interesting findings emerge. At the end of 1999 Daiwa is the only investment house preferring bonds to stocks while Julius Baer usually characterized for a strong preference for bonds behave exactly in the opposite way. This could explain the opposite performance of their portfolios.

1.6.1 Testing Homogeneity of Preferences and Beliefs

The parameters' estimates presented in the previous section are based on a single equation framework. Another possibility is to use the portfolio policy errors from all investment houses in a system. This strategy has two advantages. First, it exploits possible correlations in the portfolio weights recommended by different money managers, giving a more efficient estimation.

[25]See for example the literature regarding model uncertainty: Avramov (2002), Pettenuzzo and Timmermann (2005).

Second, we can test the restrictions that the risk aversion coefficients or the parameters driving the beliefs are the same across investment houses. To clarify the estimation process, let $e_t^{(j)}(\psi^{(j)})$ be the $m \times 1$ vector of portfolio policy error defined in (1.15) for the j-th money manager, and let $\mathbf{e}_t(\Psi) = [e_t^{(1)}(\psi^{(1)}), ..., e_t^{(J)}(\psi^{(J)})]$ be the $((m \times J) \times 1)$ vector containing the portfolio policy errors for the J investment banks participating to the poll and $\Psi = [\psi^{(1)}, ..., \psi^{(J)}]$ be the vector of parameters of interest. The GMM estimator of Ψ is given by

$$\widehat{\Psi} = arg \min_{\Psi}[g_T(\Psi)' W_T g_T(\Psi)], \tag{1.18}$$

where $g_T(\Psi) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{e}_t(\Psi)$. The estimation is done using all the available data, so we need to deal with an unbalanced panel of observations. When restrictions are considered across equations, the off diagonal blocks come into play and the elements in these blocks need to be adjusted. The parameters' estimates are not reported because in line with the ones in Tables 4 and 5. We focus on testing the restrictions of homogeneity in preferences and conditional beliefs across different money managers. The Wald test of the restrictions and their p-values are reported in Table 7. The upper panel reports the results when the restrictions are imposed across all banks, while the bottom panel displays the results for the subset of Swiss banks (Credit Suisse, Bank Julius Baer, UBS). We can always reject the restrictions of preferences and beliefs' homogeneity across all money managers. When we use the state variable proxing for the momentum effect in the portfolio policy functions, we cannot reject the null of λ_{22} being the same across banks. This implies that there is agreement among the money managers on the fact that momentum effect does not affect the beliefs on the bond market. Turning to the subset of Swiss banks, interesting findings emerge. Also in this case we find heterogeneity in the risk aversion coefficient, but we cannot reject the restrictions of the beliefs' parameters being the same conditioning on different state variables. This result imply that the parameters driving the beliefs' dynamics of the Swiss banks can be considered to be the same for the available sample.

1.6.2 The Rationality of Asset Allocation Recommendations

Recent articles on asset allocation by Canner, Mankiw, and Weil (1997) and Elton and Gruber (2000) try to develop test of investor rationality and to examine the rationality of the advice of a set of well-known advisors. The Canner, Mankiw, and Weil (1997) test of rationality for asset allocation advice is: the ratio of bonds to stocks should rise as an investor is willing to take more risk. Elton and Gruber (2000) show that the bond-stock mix can either increase or decrease as risk increases over low risk levels, but the bond-stock mix must decrease as risk increases over high risk levels. As explained in Section 4.3 in our framework the rationality test is a byproduct of the GMM estimation. Indeed Hansen's J-test tells us if the investment houses are using efficiently all the information contained in the instruments when providing their portfolio recommendations. We estimate the same model entertained in the previous section but we need to add extra variables as instruments.[26] Our new instruments' sets comprise (1) a constant, a lagged value of the state variable included in the policy function, and lagged excess returns on stocks and bonds; (2) a constant, and lagged values of three factors summarizing the state of the economy. To reduce the dimensionality of the system, we apply the methodology of dynamic factor models to a group of variables comprising real activity measures (inflation and industrial production), interest rates measures (risk-free, default spread, term spread), financial markets measures (excess returns on equity and bonds).[27] This leaves us with three variables that summarize the state of the economy. More precisely, we first normalize each series separately to have zero mean and unit variance. We then stack the seven variables into a vector z_t which can be represented as:

$$z_t = Cf_t + u_t, \tag{1.19}$$

[26] A possible concern is that of weak instruments and weak identification. According to Stock et al. (2004) if identification is weak then GMM estimates can be sensitive to the addition of instruments, so if this occurs in an empirical application it can be indicative of weak identification.

[27] This is a common practice in the literature on term structure modeling. See for example Ang and Piazzesi (2003).

where C is the factor loading matrix, f_t is the vector of factors. The error term u_t satisfies $E(u_t) = 0$ and $var(u_t) = \Gamma$, where Γ is diagonal. The extracted macro factors f_t inherit the zero mean from z_t, $(E(f_t) = 0)$ and have unit variance as any principal component $(var(f_t) = 1)$. Over 60% (75%) of the variance of variables in z_t is explained by just the first two (three) principal components. We decided to include the first three factors since the gain from the inclusion of an additional factor is limited.[28]

The results are reported in Table 8. We do not report the estimated values of the parameters since they are in general robust to different instruments sets. Turning to the J-test, we can reject rationality of the money managers in 10% and 26% of the cases at 5% level for the two sets of instruments.[29] The result is very interesting because we are able to explain the investment strategies recommended by the majority of investment banks within a simple rational model based on the Markowitz framework. Nevertheless, some caveats are necessary because the average number of observations is 59 and the power of the test could be undermined.

1.7 Conclusion

In this paper, we develop a general framework to examine the factors driving money managers' investment strategies. We treat the recommended portfolio weights as the outcome of a possibly misspecified econometric model and link it to the decision problem of the money manager. We establish conditions under which, if the money manager acts optimally, the parameters describing the preferences and conditional beliefs about the future state of the market can be recovered from the observed portfolio weights. Under the mean-variance framework we provide joint estimates of the preference parameters and the conditional market beliefs that most closely reproduce the dynamics

[28]The marginal contribution of the fourth factor to the explanation of the variance of the panel is below 10%.

[29]The percentages reported are computed across investment houses and policy functions based on different state variables.

of the observed portfolio recommendations made by a panel of international money managers for *The Economist*.

Using the empirical results reported in Section 6 we find heterogeneity in both the preferences and beliefs of our money managers. The parameters' estimates are economically reasonable and suggest that (1) money managers behave as low-risk averse investors; (2) heterogeneity in the beliefs is what really drives the portfolio recommendations; (3) the key source of disagreement is the diverse interpretation of the signals concerning the state of the economy. In particular, the banks that joined the Portfolio Poll after the nineties make little use of variables tracking the state of the economy when setting their investment strategies. We confirm the importance of standard macro (inflation and growth rate of industrial production) and financial (short term interest rate, default spread, momentum) variables as the *key drivers* of the asset allocation dynamics. We also find that under mean-variance preferences only few banks are not processing efficiently the information on the state of the economy to make their asset allocation recommendations, i.e. their investment strategies are in general rational.

The empirical findings of the present work carry practical implications for investors. An understanding of the patterns of professional advisors' investment policies is important for investors who may choose to implement a portion of their portfolio strategy using published recommendations. An attractive feature of the framework proposed in this paper is its generality which means that it could be used in future research to revisit the empirical results on mutual/pension funds performance trying to understand what drives their asset allocation.

These are two ways in which research along these lines can proceed. First, according to the literature on delegated portfolio management it would be nice to explicitly allow for relative performance concerns in the utility function of the money managers. Second, it might be interesting to use this framework to examine what drives sector rotations in the portfolios of equity mutual funds.

1.8 Appendix A

This Appendix reports the assumptions and proofs of the propositions in the paper.

1.8.1 Assumptions on the Data

Assumption 1 (A1) *The parameter space, Θ, is a compact subset of* R^k *and* θ^* *is interior to* Θ.

Assumption 2 (A2) $U(\cdot)$ *is globally concave.*

Assumption 3 (A3) $U'(\cdot)$ *is continuous and twice differentiable function with respect to* θ *in a neighborhood of* θ^*.

Assumption 4 (A4) *Returns and forecasting variables are realizations from a strictly stationary* $\mathbb{R}^m \times \mathbb{R}^k$ *-valued process;*

Assumption 5 (A5) *For each realization of the forecasting variable* X_t *there is an identifiable portfolio policy* $\omega^*(X_t, \theta^*(\psi))$ *(i.e.* $\omega^*(X_t, \theta_2^*(\psi)) = \omega^*(X_t, \theta_1^*(\psi))$ *for each realization of* X_t *implies* $\theta_1^* = \theta_2^*$*) which is a unique zero of* $E[U'(\cdot)]$.

Assumption 6 (A6) *For every* t, $E\left[J_\theta\omega\nabla_{\omega_t}W_{t+h}(\nabla_{\omega_t}W_{t+h}J_\theta\omega)'\right]$ *exists and is positive definite. If the portfolio policy function is linear the condition becomes* $E[R'_{t+h}X_tX'_tR_{t+h}]$ *exists and is positive definite.*

Assumption 7 (A7) $W_{t+h} = W(\omega_t(X_t, \theta(\psi)), R_{t+h})$ *is linear in* ω_t, *so that* $\nabla_\omega W_{t+h}$ *is independent of* ω.

1.8.2 Proofs

Proof of Proposition 1. The proof follows the line of the proof of Proposition 1 in Elliott, Komunjer, and Timmermann (2005) since our case can be seen as a multivariate extension of their problem. From assumption [A1] we know that θ^* interior to Θ is a solution to $\max_{\theta\in\Theta} EU(\theta)$ where $EU(\theta) = E[U(\theta)]$ and $U(\theta) = [U(W(\omega(X_t,\theta(\psi)), R_{t+h}), \psi)]$. Moreover, the function $U(\theta)$ is continuously differentiable on Θ. Let $\nabla_\theta EU$ be the gradient of $EU(\theta)$ on Θ. For given ψ, if $\theta^* \in \Theta$ is the max of $EU(\theta)$, then θ^* is a solution to $\nabla_\theta EU(\theta) = 0$ where $\nabla_\theta EU(\theta) = E[J_\theta\omega_t\nabla_\omega W_{t+h}\frac{\partial U(W_{t+h},\psi)}{\partial W_{t+h}}]$, $\nabla_{\omega_t} W_{t+h} = [R_{1,t+h},\ldots,R_{m,t+h}]' = R'_{t+h}$ and $J_\theta\omega_t = \begin{pmatrix} \frac{\partial\omega_{1,t}}{\partial\theta_1} & \cdots & \frac{\partial\omega_{1t}}{\partial\theta_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial\omega_{m,t}}{\partial\theta_1} & \cdots & \frac{\partial\omega_{m,t}}{\partial\theta_m}\end{pmatrix}$, and if the portfolio policy is linear $J_\theta\omega_t = \begin{pmatrix} X_t & \cdots & \cdots \\ \vdots & \ddots & \vdots \\ \cdots & \cdots & X_t\end{pmatrix}$. We show that (1.2) holds which completes the necessary part of the proof. Let $H(\theta)$ be the hessian matrix of $EU(\theta)$ with respect to θ. We know that θ^* is a local maxima of θ if $\nabla_\theta EU(\theta^*) = 0$ and $H(\theta^*)$ is negative definite. The first order condition $\nabla_\theta EU(\theta^*) = 0$ is implied by (1.2). We show that $H(\theta^*)$ is negative definite. Under assumption [A6-A7] we have that

$$H(\theta) = E\begin{bmatrix} U(\cdot)''J_\theta\omega_t\nabla_\omega W_{t+h}(\nabla_\omega W_{t+h}J_\theta\omega_t)' & 0 \\ 0 & U(\cdot)''J_\theta\omega_t\nabla_\omega W_{t+h}(\nabla_\omega W_{t+h}J_\theta\omega_t)'\end{bmatrix} \tag{1.20}$$

and if the portfolio policy is linear, we have that

$$H(\theta) = E\begin{bmatrix} U(\cdot)''R'_{t+h}X_tX'_tR_{t+h} & 0 \\ 0 & U(\cdot)''R'_{t+h}X_tX'_tR_{t+h}\end{bmatrix} \tag{1.20}$$

Assumption [A2] guarantees that $U(\cdot)'' < 0$ and assumption [A6] guarantees that $E\left[J_\theta\omega\nabla_{\omega_t}W_{t+h}(\nabla_{\omega_t}W_{t+h}J_\theta\omega)'\right]$ exists and positive definite. Thus the elements of the main diagonal are negative. The matrix $H(\theta)$ is negative definite if all the eigenvalues are negatives.

Since $H(\theta)$ is diagonal, the eigenvalues are the elements on the main diagonal which we have shown to be negative. So for every $\theta \in \Theta$ the matrix $H(\theta)$ is negative definite, then so must be for θ^*. Thus any $\omega_t = \omega(X_t, \theta(\psi))$ which satisfies the moment condition (1.2) is a solution to the asset allocation problem.

We now use the implicit function theorem to show that for any realization of X_t, the function $\omega_t^* = \omega(X_t, \theta^*(\psi))$ defined implicitly by (1.2) is a one-to-one mapping from the set of parameters ψ_0 to the set of portfolio weights, ω_t. Define $\xi(\psi, \theta) \equiv \nabla_\theta EU(\theta^*)$, so that $\xi(\psi, \theta^*) = 0$, by (1.2). Furthermore, the function is continuously differentiable and we have that $\frac{\partial \xi(\psi,\theta)}{\partial \theta} = H(\theta)$, and $\frac{\partial \xi(\psi,\theta)}{\partial \psi} = U(\cdot)'' \frac{\partial W(\cdot)}{\partial \psi}$. Finally we have that the matrix $\frac{\partial \xi(\psi,\theta^*)}{\partial \theta}$ is non singular given that $H(\theta)$ is negative definite. We can now apply the implicit function theorem to show that for every ψ there exists a neighborhood D of ψ and a neighborhood V of θ^* such that the system of equations $\xi(\psi, \theta) = 0$ has a unique solution θ^* in V, and the function $\theta = \theta(\psi)$ defined implicitly by $\xi(\psi, \theta) = 0$ is continuously differentiable from D to V. In particular we have that, for any realization of X_t, the function $\omega_t^* = \theta^*(\psi) X_t$ is continuously differentiable from $\mathcal{G} \subseteq \mathrm{R}^p$ to $A \subseteq \mathrm{R}^m$. We need to show that $\theta^*(\psi) X_t$ is a one-to-one mapping from $\mathcal{G}$ to A. It is surjective by construction, so we need to show that it is injective on $\mathcal{G}$, i.e. $\theta^*(\psi_1) = \theta^*(\psi_2)$ implies $\psi_1 = \psi_2$. Using identifiability of a linear portfolio policy implied by [A5], we know that for each realization of X_t there is a unique $\theta^* \in \Theta$ such that $\omega^* = \theta^* X_t$ so by using the previous result there is a unique ψ such that $\omega^* = \theta^*(\psi) X_t$.■

Bibliography

Admati, Anar A., and Paul C. Pfeiderer, 1997, Does it all add up? benchmarks and the compensation of active portfolio managers, *Journal of Business* 70, 323–350.

Ait-Sahalia, Yacine, and Micheal Brandt, 2001, Variable selection for portfolio choice, *Journal of Finance* 56, 1297–1351.

Ang, Andrew, and Monika Piazzesi, 2003, A no-arbitrage vector-autoregression of term structure dynamics with macroeconomic and latent variables, *Journal of Monetary Economics* 50, 745–787.

Annaert, Jan, Mark J.K. De Ceuster, and Wim Van Hyfte, 2004, The value of asset allocation advice, *Journal of Money Credit and Banking* 29, 661–680.

Arshanapalli, Bala, Daniel T. Coggin, and William Nelson, 2001, Is fixed-weight asset allocation really better?, *Journal of Portfolio Management* 27, 27–34.

Avramov, Doron, 2002, Stock return predictability and model uncertainty, *Journal of Financial Economics* 64, 423–458.

Bange, Mary, Kenneth Khang, and Thomas W. Miller, 2004, Return momentum and global portfolio allocations, *Journal of Empirical Finance* 11, 429–459.

Barber, Brad M., and Douglas Loeffler, 1993, The dartboard column: Second-hand information and price pressure, *Journal of Financial and Quantitative Analysis* 28, 273–284.

Basak, Suleyman, Anna Pavlova, and Alex Shapiro, 2004, Offsetting the incentives: risk shifting and benefits of benchmarking in money management, NBER Working Paper.

Becker, Connie, Wayne Ferson, David Myers, and Michael Schill, 1999, Conditional market timing with benchmark investors, *Journal of Financial Economics* 52, 119–148.

Blake, David, Bruce N. Lehmann, and Allan Timmermann, 1999, Asset allocation dynamics and pension funds performance, *Journal of Business* 72, 429–462.

Bliss, Robert, and Nikolaos Panigirtzoglou, 2004, Option-implied risk aversion estimates, *Journal of Finance* 59, 407–446.

Brandt, Micheal, 1999, Estimating portfolio and consumption choice: A conditional euler equations approach, *Journal of Finance* 54, 1609–1646.

——— , 2005, Portfolio choice problems, forthcoming in Y. Ait-Sahalia and L.P. Hansen, eds., Handbook of Financial Econometrics, Elsevier Science.

——— , and Pedro Santa-Clara, 2004, Dynamic portfolio selection by augmenting the asset space, forthcoming Journal of Finance.

——— , and Rossen Valkanov, 2005, Equity portfolios with parametric weights: Exploiting size, book to market, and momentum, Working paper, Duke University.

Brennan, Micheal J., 1993, Agency and asset pricing, working paper, UCLA.

Campbell, John Y., 1987, Stock returns and the term structure, *Journal of Financial Economics* 18, 377–399.

——— , and Luis M. Viceira, 2002, *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors* (Oxford University Press).

Canner, Niko, N. Gregory Mankiw, and David N. Weil, 1997, An asset allocation puzzle, *American Economic Review* 87, 181–191.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Chevalier, Judith, and Glenn Ellison, 1997, Risk taking by mutual funds as a response to incentives, *Journal of Political Economy* 105, 1167–1200.

Cuoco, Domenico, and Robert Kaniel, 2000, General equilibrium implications of fund managers compensation fees, working paper, University of Pennsylvania and University of Texas at Austin.

Cuoco, Domenico, and Fernando Zapatero, 2000, On the recoverability of preferences and beliefs, *Review of Financial Studies* 13, 417–431.

Dybvig, Philip, and Hector Polemarchakis, 1981, Recovering cardinal utility, *Review of Economic Studies* 48, 159–166.

Elliott, Graham, Ivana Komunjer, and Allan Timmermann, 2005, Estimation and testing of forecast rationality under flexible loss, *Review of Economic Studies* 72, 1107–1125.

Elton, Elton J., and Martin J. Gruber, 2000, The rationality of asset allocation recommendations, *Journal of Financial and Quantitative Analysis* 36, 27–41.

Fama, Eugene F., and Kenneth R. French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 2349.

Farnsworth, Heber, and Jonathan Taylor, 2004, Evidence on the compensation of portfolio managers, working paper, Olin School of Business at Washington University.

Ferson, Wayne E., and Campbell R. Harvey, 1991, The variation of economic risk premiums, *Journal of Political Economy* 99, 385–415.

Foster, F. Douglas, and Michael Stutzer, 2003, Performance and risk aversion of funds with benchmarks: a large deviation approach, working paper.

French, Kenneth, and James M. Poterba, 1991, International diversification and international equity markets, *American Economic Review, Papers and Proceedings.*

Goetzmann, William, Jonathan Ingersoll, and Stephen A. Ross, 2003, High-water marks and hedge fund management contracts, *Journal of Finance* 58, 16851717.

Graham, John R., 1999, Herding among investment newsletters: Theory and evidence, *Journal of Finance* 54, 237–268.

———, and Campbell R. Harvey, 1996, Market timing ability and volatility implied in investment newsletters' asset allocation recommendations, *Journal of Financial Economics* 42, 97–421.

Grinblatt, Mark, and Sheridan Titman, 1993, Performance measurement without benchmarks: An examination of mutual fund returns, *Journal of Business* 66, 47–68.

Hansen, Lars Peter, and Kenneth J. Singleton, 1982, Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica* 50, 1269–1286.

Huddart, Stephen, 1999, Reputation and performance fee effects on portfolio choice by investment advisers, *Journal of Financial Markets* 2, 227–271.

Jaffe, Jeffrey F., and James M. Mahoney, 1999, The performance of investment newsletters, *Journal of Financial Economics* 53, 289–307.

Kapur, Sandeep, and Allan Timmermann, 2004, Relative performance evaluation contracts and asset market equilibrium, Forthcoming Economic Journal.

Kraus, Alan, and Gordon A. Sick, 1980, Distinguishing beliefs and preferences in equilibrium prices, *Journal of Finance* 35, 335–344.

Lamont, Owen, 2001, Economic tracking portfolios, *Journal of Econometrics* 105, 161–184.

Lehmann, Bruce N., 2005, The role of beliefs in inference for rational expectations models, working paper, UC San Diego.

Pesaran, Hashem, and Allan Timmermann, 1995, Predictability of stock returns: Robustness and economic significance, *Journal of Finance* 50, 1201–1228.

Wolf, Charles, and Larry Pohlman, 1983, The recovery of risk preferences from actual choices, *Econometrica* 51, 843-850.

Table 1.1: Recommended Asset Allocation: Summary Statistics

The table reports the number of recommendations (T) for each investment house and the minimum, average and maximum percentage allocations into equity, bonds, and cash. In parentheses we report the nationality of each investment bank.

		Equity					Bond					Cash				
	T	Min	Mean	Max	Std	AR(1)	Min	Mean	Max	Std	AR(1)	Min	Mean	Max	Std	AR(1)
Commerz Internat.(EU)	45	0.35	0.55	0.94	0.09	0.95	0.06	0.42	0.50	0.07	0.98	0.00	0.03	0.25	0.06	0.67
Credit Suisse (EU)	48	0.25	0.37	0.50	0.06	0.97	0.31	0.45	0.64	0.08	0.98	0.03	0.18	0.34	0.10	0.94
Credit Agricole (EU)	22	0.40	0.57	0.70	0.08	0.95	0.25	0.35	0.50	0.06	0.95	0.00	0.08	0.20	0.06	0.86
Bank Julius Baer (EU)	91	0.32	0.43	0.55	0.06	0.98	0.37	0.47	0.62	0.06	0.99	0.00	0.10	0.21	0.05	0.92
Robeco Group (EU)	47	0.40	0.51	0.60	0.05	0.98	0.40	0.48	0.60	0.05	0.98	-0.03	0.02	0.10	0.03	0.53
Merrill Lynch (US)	26	0.38	0.49	0.60	0.08	0.97	0.30	0.42	0.55	0.08	0.94	0.00	0.09	0.15	0.05	0.85
Lehman Brothers (US)	46	0.47	0.62	0.86	0.10	0.97	0.11	0.32	0.45	0.08	0.96	0.00	0.05	0.21	0.07	0.80
Brown Brothers Harriman (US)	66	0.45	0.78	0.91	0.16	0.99	0.00	0.08	0.48	0.14	0.77	0.06	0.14	0.30	0.06	0.96
Scudder Stevens (US)	66	0.59	0.85	1.00	0.10	0.99	0.00	0.09	0.21	0.06	0.95	0.00	0.07	0.20	0.05	0.87
Standard Life (UK)	20	0.55	0.57	0.62	0.03	0.95	0.37	0.42	0.45	0.03	0.95	0.00	0.01	0.08	0.02	0.00
Capital House (UK)	66	0.20	0.66	0.90	0.16	0.98	0.09	0.20	0.50	0.10	0.95	0.00	0.14	0.50	0.11	0.86
UBS/Phillips Drew (EU)	89	0.35	0.52	0.70	0.11	0.99	0.23	0.42	0.57	0.09	0.98	0.00	0.06	0.20	0.06	0.97
Daiwa (JP)	106	0.20	0.64	0.90	0.15	0.99	0.10	0.30	0.70	0.14	0.95	0.00	0.06	0.45	0.08	0.76
Nikko (JP)	22	0.55	0.66	0.80	0.08	0.95	0.15	0.27	0.35	0.06	0.96	0.00	0.07	0.15	0.04	0.82
Consensus	59	0.39	0.60	0.79	0.14	0.97	0.14	0.31	0.50	0.13	0.95	0.01	0.09	0.20	0.07	0.79

Table 1.2: Portfolio Composition

The table reports the optimal mix of the risk-free asset and the optimal risky portfolio (Cash/(Equity+Bonds)) and optimal mix of the risky portfolio (Bonds/Equity).

	Cash/(Equity+Bonds)				Bonds/Equity			
	Min	Mean	Max	Std	Min	Mean	Max	Std
Commerz Internat.	0.00	0.04	0.32	0.07	0.07	0.80	1.24	0.19
Credit Suisse	0.03	0.24	0.52	0.15	0.80	1.28	2.29	0.35
Credit Agricole	0.00	0.09	0.25	0.07	0.36	0.63	1.25	0.20
Bank Julius Baer	0.00	0.11	0.27	0.06	0.70	1.11	1.86	0.27
Robeco Group	-0.03	0.02	0.11	0.03	0.67	0.95	1.50	0.18
Merrill Lynch	0.00	0.10	0.18	0.05	0.50	0.90	1.38	0.30
Lehman Brothers	0.00	0.06	0.27	0.08	0.13	0.55	0.90	0.20
Brown Brothers Harriman	0.06	0.17	0.43	0.09	0.00	0.16	1.04	0.30
Scudders Stevens	0.00	0.07	0.26	0.06	0.00	0.11	0.36	0.09
Standard Life	0.00	0.01	0.09	0.02	0.61	0.74	0.82	0.08
Capital House	0.00	0.18	1.00	0.19	0.10	0.37	1.50	0.32
UBS/Phillips Drew	0.00	0.07	0.25	0.07	0.34	0.87	1.51	0.37
Daiwa	0.00	0.08	0.82	0.12	0.11	0.57	2.33	0.51
Nikko	0.00	0.08	0.18	0.05	0.19	0.42	0.58	0.12

Table 1.3: Portfolio Performance

The table reports the results for the asset allocation choices among equities, bonds, and cash for all investment houses. 60/30/10 is the mean return differential with respect to the static 60/30/10 allocation. Consensus is the mean return differential with respect to the average performance generated by the competitors. We use quarterly returns in percentage. PCM is the Portfolio Change Measure of Grinblatt and Titman. LMM is the Lagged Momentum Measure of Grinblatt, Titman, and Wermers. Panel A reports estimates for the full sample, while panel B focuses on the investment houses active before the October 1987 Crash. Standard errors are reported in parentheses. ** and * denote significance at 5% and 10% level respectively.

Panel A: Full sample: 1981-2004

	Performance		PCM			LMM		
	60/30/10	Consensus	Portfolio	Equity	Bond	Portfolio	Equity	Bond
Commerz Internat.	0.10	0.09	-0.01	-0.00	-0.01	-0.00	-0.03	0.03
	(0.09)	(0.07)	(0.05)	(0.03)	(0.04)	(0.05)	(0.05)	(0.03)
Credit Suisse	-0.19	-0.19	0.00	-0.00	0.00	0.09	0.06	0.02
	(0.24)	(0.18)	(0.04)	(0.05)	(0.02)	(0.07)	(0.06)	(0.02)
Credit Agricole	-0.16	0.02	0.01	0.02	-0.01	0.17**	0.12*	0.04
	(0.12)	(0.11)	(0.10)	(0.10)	(0.02)	(0.08)	(0.07)	(0.04)
Bank Julius Baer	-0.26*	-0.43*	0.10**	0.07**	0.03*	0.04	0.03	0.01
	(0.15)	(0.23)	(0.04)	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)
Robeco Group	0.04	0.07	0.03	0.01	0.01	-0.04	-0.03	0.00
	(0.12)	(0.07)	(0.03)	(0.03)	(0.01)	(0.03)	(0.03)	(0.02)
Merrill Lynch	-0.27	-0.14	0.05	0.12**	-0.06	-0.01	-0.02	0.02
	(0.17)	(0.12)	(0.06)	(0.06)	(0.06)	(0.09)	(0.08)	(0.06)
Lehman Brothers	0.04	0.03	-0.03	0.01	-0.03	-0.07	-0.04	-0.02
	(0.10)	(0.16)	(0.07)	(0.07)	(0.03)	(0.06)	(0.05)	(0.03)
Brown Brothers Harriman	0.39	0.20	0.12	0.08	0.05	0.14	0.17	-0.02
	(0.26)	(0.21)	(0.15)	(0.12)	(0.08)	(0.18)	(0.16)	(0.04)
Scudders Stevens	0.45	0.29	0.01	0.02	-0.01	0.01	0.03	-0.01
	(0.30)	(0.26)	(0.06)	(0.06)	(0.01)	(0.04)	(0.05)	(0.02)
Standard Life	0.13	-0.06	-0.00	-0.00	-0.00	0.05	0.06	-0.01
	(0.15)	(0.12)	(0.03)	(0.02)	(0.01)	(0.04)	(0.05)	(0.02)
Capital House	0.21	0.00	0.23**	0.23**	0.00	0.11*	0.07	0.04
	(0.19)	(0.14)	(0.08)	(0.11)	(0.04)	(0.07)	(0.08)	(0.05)
UBS/Phillips Drew	-0.03	-0.21	-0.00	0.00	-0.00	-0.03	-0.03	0.00
	(0.10)	(0.17)	(0.05)	(0.05)	(0.04)	(0.05)	(0.05)	(0.04)
Daiwa	0.38**	0.28**	0.16	0.22**	-0.06	-0.13	-0.10	-0.03
	(0.13)	(0.11)	(0.12)	(0.11)	(0.06)	(0.09)	(0.08)	(0.05)
Nikko	0.18	0.40	0.17**	0.14*	0.03	0.03	0.06	-0.03
	(0.16)	(0.27)	(0.06)	(0.08)	(0.03)	(0.04)	(0.05)	(0.04)

Panel B: Before October 1987

	Performance		PCM			LMM		
	60/30/10	Consensus	Portfolio	Equity	Bond	Portfolio	Equity	Bond
Bank Julius Baer	-0.53**	-1.07**	0.11**	0.07**	0.03	0.08*	0.06	0.02
	(0.26)	(0.46)	(0.05)	(0.03)	(0.03)	(0.05)	(0.04)	(0.05)
Brown Brothers Harriman	0.63**	0.19	0.16**	0.16**	0.00	0.12**	0.14**	-0.02
	(0.27)	(0.23)	(0.06)	(0.07)	(0.02)	(0.05)	(0.07)	(0.02)
Scudders Stevens	0.76**	0.34	0.10	0.12*	-0.01	0.05	0.07	-0.01
	(0.28)	(0.25)	(0.07)	(0.07)	(0.02)	(0.06)	(0.07)	(0.03)
Capital House	0.40*	-0.08	0.31**	0.31*	0.01	0.17*	0.12	0.06
	(0.21)	(0.19)	(0.12)	(0.17)	(0.07)	(0.10)	(0.12)	(0.08)
UBS	0.23*	-0.34	-0.13**	-0.08	-0.06	-0.13**	-0.09	-0.04
	(0.12)	(0.22)	(0.06)	(0.07)	(0.06)	(0.06)	(0.08)	(0.08)
Daiwa	1.04**	0.71**	-0.08	0.09	-0.17	-0.11	0.03	-0.14
	(0.24)	(0.22)	(0.08)	(0.06)	(0.13)	(0.10)	(0.02)	(0.12)

Table 1.4: Risk Aversion Estimates

The table reports iterated GMM estimates with 2 equations and 2 instruments (constant and lagged state variable). The 2 equations are derived from the expression of $\omega^*_{1,t}$ and $\omega^*_{2,t}$. The dynamics of the beliefs in the Equity and Bond Markets are given by $s_{i,t+1} = \lambda_{i1} + \lambda_{i2} z_t + \varepsilon_{i,t}, \quad i = 1, 2$. We report results for policy functions based on different z_t. The value of λ_{11} has been normalized to 1. We standardize the conditioning variables to ease the interpretation and comparison of the portfolio policy functions. ** and * denote significance at 5% and 10% level respectively. Standard errors in parentheses.

	G7 Inflation	G7 Ind. Prod.	Risk-free Rate	Default Spread	Term Spread	Return on Equity-Bond
Commerz Intern.	1.70**	1.87**	1.86**	1.88**	1.85**	1.85**
	(0.14)	(0.06)	(0.06)	(0.10)	(0.05)	(0.05)
Credit Suisse	2.63**	2.73**	2.71**	2.74**	2.70**	2.72**
	(0.15)	(0.11)	(0.11)	(0.16)	(0.11)	(0.13)
Credit Agricole	1.70**	1.77**	1.86**	2.56**	1.77**	1.62**
	(0.14)	(0.07)	(0.08)	(0.90)	(0.09)	(0.26)
Bank Julius Baer	2.40**	2.31**	2.31**	2.38**	2.31**	2.32**
	(0.07)	(0.06)	(0.06)	(0.05)	(0.07)	(0.06)
Robeco Group	1.92**	1.96**	1.97**	1.96**	1.97**	1.96**
	(0.07)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)
Merrill Lynch	2.22**	1.99**	1.94**	1.84**	2.03**	2.06**
	(0.41)	(0.09)	(0.11)	(0.34)	(0.08)	(0.21)
Lehman Brothers	1.51**	1.62**	1.59**	1.47**	1.63**	1.63**
	(0.12)	(0.07)	(0.06)	(0.04)	(0.07)	(0.08)
Brown Brothers Harriman	1.20**	1.28**	1.26**	1.21**	1.28**	1.29**
	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)
Scudder Stevens	1.12**	1.18**	1.17**	1.13**	1.17**	1.18**
	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Capital House	1.40**	1.52**	1.49**	1.40**	1.52**	1.53**
	(0.07)	(0.06)	(0.06)	(0.05)	(0.08)	(0.07)
UBS/Phillips Drew	1.97**	1.91**	1.91**	1.96**	1.92**	1.91**
	(0.09)	(0.08)	(0.08)	(0.06)	(0.08)	(0.08)
Daiwa Europe	1.54**	1.55**	1.55**	1.55**	1.55**	1.55**
	(0.06)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)
Nikko Securities	1.42**	1.53**	1.51**	2.56*	1.46**	1.34**
	(0.11)	(0.04)	(0.09)	(1.32)	(0.05)	(0.32)

Table 1.5: Conditional Beliefs' Dynamics

The table reports iterated GMM estimates with 2 equations and 2 instruments (constant and lagged state variable). The 2 equations are derived from the expression of $\omega^*_{1,t}$ and $\omega^*_{2,t}$. The dynamics of the beliefs in the Equity and Bond Markets are given by $s_{i,t+1} = \lambda_{i1} + \lambda_{i2} z_t + \varepsilon_{i,t}, \quad i = 1, 2$. We report results for policy functions based on different z_t. The value of λ_{11} has been normalized to 1. We standardize the conditioning variables to ease the interpretation and comparison of the portfolio policy functions. ** and * denote significance at 5% and 10% level respectively. Standard errors in parentheses.

	G7 Inflation			G7 Ind. Prod.			Risk-free Rate		
	λ_{12}	λ_{21}	λ_{22}	λ_{12}	λ_{21}	λ_{22}	λ_{12}	λ_{21}	λ_{22}
Commerz Intern.	0.17	0.73**	0.00	-0.03	0.80**	0.05 *	-0.01	0.80**	0.03 **
	(0.18)	(0.08)	(0.03)	(0.03)	(0.04)	(0.03)	(0.05)	(0.03)	(0.01)
Credit Suisse	0.06	1.09**	-0.19	0.04	1.24**	-0.03	-0.04	1.22**	0.06
	(0.05)	(0.11)	(0.13)	(0.04)	(0.08)	(0.08)	(0.04)	(0.08)	(0.06)
Credit Agricole	0.06	0.64**	0.09	0.06	0.61**	0.07 *	0.18 **	0.65**	0.02
	(0.11)	(0.11)	(0.15)	(0.07)	(0.06)	(0.04)	(0.04)	(0.07)	(0.04)
Bank Julius Baer	-0.20**	1.16**	0.17 **	-0.02	1.08**	0.00	-0.02	1.09**	-0.04
	(0.10)	(0.05)	(0.07)	(0.03)	(0.05)	(0.03)	(0.03)	(0.05)	(0.03)
Robeco Group	0.04	0.92**	0.02	-0.02	0.93**	-0.00	0.04	0.94**	-0.04
	(0.05)	(0.06)	(0.03)	(0.03)	(0.04)	(0.02)	(0.04)	(0.04)	(0.03)
Merrill Lynch	-0.37	0.97**	0.14	-0.08	0.83**	0.03	-0.08	0.83**	-0.04
	(0.59)	(0.26)	(0.35)	(0.07)	(0.09)	(0.09)	(0.08)	(0.10)	(0.06)
Lehman Brothers	0.12	0.50**	0.01	-0.06	0.53**	0.05	-0.10**	0.51**	0.07 **
	(0.11)	(0.07)	(0.05)	(0.05)	(0.06)	(0.04)	(0.04)	(0.05)	(0.03)
Brown Brothers Harriman	-0.14**	0.06	0.08 **	0.13 **	0.10**	-0.11**	0.07 **	0.09**	-0.06**
	(0.04)	(0.04)	(0.02)	(0.05)	(0.04)	(0.04)	(0.03)	(0.04)	(0.02)
Scudder Stevens	-0.10**	0.07**	0.05 **	0.09 **	0.10**	-0.06**	0.06 **	0.09**	-0.04**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
Capital House	-0.18**	0.26**	0.05 *	0.15 **	0.31**	-0.08**	0.11 **	0.29**	-0.08**
	(0.05)	(0.05)	(0.03)	(0.05)	(0.04)	(0.02)	(0.04)	(0.03)	(0.03)
UBS/Phillips Drew	0.13 **	0.83**	-0.08	-0.04	0.79**	0.02	-0.09**	0.79**	0.04
	(0.06)	(0.07)	(0.06)	(0.06)	(0.07)	(0.04)	(0.04)	(0.06)	(0.03)
Daiwa Europe	-0.08	0.45**	0.09	0.12 **	0.45**	-0.14**	0.05	0.45**	-0.10**
	(0.07)	(0.05)	(0.08)	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)
Nikko Securities	0.11	0.38**	0.01	0.10	0.42**	-0.05	0.02	0.39**	0.04
	(0.13)	(0.08)	(0.06)	(0.07)	(0.03)	(0.06)	(0.08)	(0.06)	(0.05)

	Default Spread			Term Spread			Return on Equity-Bond		
	λ_{12}	λ_{21}	λ_{22}	λ_{12}	λ_{21}	λ_{22}	λ_{12}	λ_{21}	λ_{22}
Commerz Intern.	0.02	0.78**	0.04	-0.02	0.80**	-0.03**	0.07	0.79**	-0.02
	(0.06)	(0.05)	(0.03)	(0.03)	(0.03)	(0.01)	(0.10)	(0.03)	(0.05)
Credit Suisse	0.01	1.03**	0.32 **	-0.07**	1.23**	-0.04	0.22	1.24**	0.14
	(0.09)	(0.08)	(0.07)	(0.03)	(0.08)	(0.06)	(0.17)	(0.09)	(0.16)
Credit Agricole	0.49	0.41	0.52 **	0.05	0.64**	-0.08**	-0.38	0.53**	0.22
	(0.55)	(0.38)	(0.20)	(0.06)	(0.06)	(0.03)	(0.79)	(0.15)	(0.22)
Bank Julius Baer	0.12 **	1.14**	-0.09**	-0.03	1.08**	0.02	0.04	1.09**	-0.02
	(0.03)	(0.04)	(0.03)	(0.03)	(0.05)	(0.03)	(0.04)	(0.05)	(0.04)
Robeco Group	0.00	0.96**	-0.05	0.02	0.94**	-0.03	-0.07	0.93**	0.02
	(0.05)	(0.06)	(0.05)	(0.02)	(0.04)	(0.02)	(0.06)	(0.04)	(0.06)
Merrill Lynch	-0.09	0.73**	0.04	0.10 **	0.86**	-0.07*	0.23	0.87**	-0.11
	(0.19)	(0.25)	(0.17)	(0.03)	(0.08)	(0.04)	(0.61)	(0.19)	(0.51)
Lehman Brothers	-0.14**	0.47**	0.02	0.02	0.54**	-0.04	-0.10	0.54**	0.12
	(0.06)	(0.03)	(0.05)	(0.04)	(0.06)	(0.03)	(0.10)	(0.07)	(0.10)
Brown Brothers Harriman	0.10 **	0.06	-0.05**	0.01	0.10**	0.00	0.19 **	0.11**	-0.11**
	(0.03)	(0.05)	(0.03)	(0.05)	(0.04)	(0.04)	(0.06)	(0.04)	(0.05)
Scudder Stevens	0.08 **	0.07**	-0.05**	0.00	0.10**	0.00	0.08 *	0.10**	-0.05*
	(0.01)	(0.01)	(0.01)	(0.03)	(0.02)	(0.02)	(0.05)	(0.02)	(0.03)
Capital House	0.14 **	0.25**	-0.06**	-0.07	0.31**	0.07 *	0.22 **	0.31**	-0.07**
	(0.03)	(0.04)	(0.01)	(0.05)	(0.05)	(0.04)	(0.07)	(0.05)	(0.04)
UBS/Phillips Drew	-0.15**	0.82**	0.08 **	0.05	0.79**	0.01	-0.04	0.79**	0.01
	(0.03)	(0.05)	(0.03)	(0.05)	(0.07)	(0.04)	(0.05)	(0.07)	(0.04)
Daiwa Europe	-0.01	0.45**	-0.04	0.05	0.45**	0.00	0.19 **	0.45**	-0.16**
	(0.06)	(0.05)	(0.05)	(0.04)	(0.06)	(0.04)	(0.05)	(0.05)	(0.07)
Nikko Securities	0.74	0.85	-0.18	-0.09**	0.38**	0.04	-0.54	0.28	0.38
	(0.95)	(0.76)	(0.48)	(0.03)	(0.05)	(0.03)	(1.17)	(0.23)	(0.75)

Table continued from previous page.

Table 1.6: Summary

The table reports the sign (+ or $-$) of λ_{i2} if significant in the policy function. If the state variable is not significant we report o. The specification of the beliefs' dynamics is given by $s_{i,t+1} = \lambda_{i1} + \lambda_{i2} z_t + \varepsilon_{i,t}, \quad i = 1,2$

	G7 Inflation		G7 Ind. Prod.		Risk-free Rate		Default Spread		Term Spread		Return on Equity-Bond	
	Equity	Bond	Equity	Bond	Equity	Bond	Equity	Bond	Equity	Bond	Equity	Bond
Commerz Intern.	o	o	o	+	o	+	o	o	o	-	o	o
Credit Suisse	o	o	o	+	o	o	o	+	-	o	o	o
Credit Agricole	o	o	o	o	+	o	o	o	o	-	o	+
Bank Julius Baer	-	+	o	o	o	o	+	-	o	o	o	o
Robeco Group	o	o	o	o	o	o	o	o	o	-	o	o
Merrill Lynch	o	o	o	o	o	o	o	o	+	-	o	o
Lehman Brothers	o	o	o	o	-	+	-	o	o	-	o	o
Brown Brothers Harriman	-	+	+	-	+	-	+	-	o	o	+	-
Scudder Stevens	-	+	+	-	+	-	+	-	o	o	+	-
Capital House	-	+	+	-	+	-	+	-	-	+	+	-
UBS/Phillips Drew	+	-	o	o	-	+	-	+	o	o	o	o
Daiwa Europe	o	+	+	-	+	-	o	o	o	o	+	-
Nikko Securities	o	o	o	o	o	o	o	o	+	o	o	o

Table 1.7: Testing Preferences and Beliefs Homogeneity across Money Managers

The table reports the Wald test based on estimates obtained by iterated multiple equations GMM. The system consists of 52 equations: 2 moment conditions for 13 money managers and 2 instruments (constant and lagged state variable). γ is the risk aversion coefficient. The dynamics of the beliefs in the Equity and Bond Markets are given by $s_{i,t+1} = \lambda_{i1} + \lambda_{i2} z_t + \varepsilon_{i,t}, \quad i = 1, 2$. We report results for policy functions based on different z_t. P-values are in parentheses.

All Banks

H_0	G7 Inflation	G7 Ind. Prod.	Risk-free Rate	Default Spread	Term Spread	Return on Equity-Bond
$\gamma^{(1)} = ... = \gamma^{(J)}$	1327 (0.00)	3027 (0.00)	2972 (0.00)	2824 (0.00)	2891 (0.00)	2409 (0.00)
$\lambda_{12}^{(1)} = ... = \lambda_{12}^{(J)}$	37.40 (0.00)	58.25 (0.00)	84.41 (0.00)	199.70 (0.00)	75.77 (0.00)	24.73 (0.00)
$\lambda_{21}^{(1)} = ... = \lambda_{21}^{(J)}$	1857 (0.00)	4663 (0.00)	3858 (0.00)	3413 (0.00)	5040 (0.00)	3787 (0.00)
$\lambda_{22}^{(1)} = ... = \lambda_{22}^{(J)}$	30.47 (0.00)	58.29 (0.00)	59.61 (0.00)	158.63 (0.00)	41.94 (0.00)	11.50 (0.40)

Swiss Banks: Credit Suisse, Bank Julius Baer, UBS Philips and Drew

H_0	G7 Inflation	G7 Ind. Prod.	Risk-free Rate	Default Spread	Term Spread	Return on Equity-Bond
$\gamma^{(1)} = ... = \gamma^{(J)}$	8.36 (0.01)	8.36 (0.01)	72.39 (0.00)	116.30 (0.00)	55.09 (0.00)	44.22 (0.00)
$\lambda_{12}^{(1)} = ... = \lambda_{12}^{(J)}$	2.03 (0.36)	2.03 (0.36)	3.92 (0.14)	118.60 (0.00)	6.80 (0.03)	3.95 (0.14)
$\lambda_{21}^{(1)} = ... = \lambda_{21}^{(J)}$	3.16 (0.21)	3.16 (0.20)	35.03 (0.00)	70.62 (0.00)	35.32 (0.00)	30.30 (0.00)
$\lambda_{22}^{(1)} = ... = \lambda_{22}^{(J)}$	3.50 (0.17)	3.50 (0.17)	4.47 (0.10)	60.32 (0.00)	2.89 (0.24)	0.62 (0.73)

Table 1.8: Rationality Test of the Money Managers' Investment Strategies

The table reports he J-test values obtained by iterated GMM estimates with over-identifying restrictions given by different sets of instruments. The first instruments' set (1) comprises a constant, a lagged value of the state variable used in the policy function, and the lagged values of excess returns on the stock and bond market. The second instruments' set (2) comprises a constant, one lagged value of three factors extracted from a group of real activity measures, interest rates measures, financial markets measures. ** and * denote significance at 5% and 10% level respectively. The p-values are computed by using a χ^2 distribution with 4 degrees of freedom.

	G7 Inflation		G7 Ind. Prod.		Risk-free Rate		Default Spread		Term Spread		Return on Equity-Bond	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Commerz Intern.	1.82	10.38**	7.77	7.84 *	4.81	6.97	2.43	4.83	3.12	6.67	3.12	8.80 *
Credit Suisse	6.59	4.32	7.64	7.09	5.69	7.78	6.33	6.03	4.51	1.22	4.51	3.85
Credit Agricole	1.78	5.88	2.94	4.47	3.51	9.77 **	2.83	8.88 *	1.86	7.40	1.86	9.42 *
Bank Julius Baer	21.84**	14.14**	7.30	16.15**	3.59	15.65**	5.37	3.88	6.28	12.64**	6.28	3.13
Robeco Group	2.83	4.06	1.28	4.50	8.28 *	4.82	1.45	13.94 **	2.82	6.76	2.82	4.48
Merrill Lynch	2.10	12.29**	1.26	6.89	1.41	6.43	1.53	7.01	1.06	2.80	1.06	11.33**
Lehman Brothers	2.14	3.60	0.66	5.36	2.59	4.26	5.69	4.04	2.19	5.53	2.19	2.94
Brown Brothers Harriman	8.38 *	4.98	3.88	3.51	8.57 *	6.06	8.11 *	6.93	6.06	7.12	6.06	5.09
Scudder Stevens	21.04**	9.29 *	6.88	15.30**	3.14	6.33	5.75	11.30 **	6.05	16.53**	6.05	6.36
Capital House	9.56 **	9.15 *	16.01**	7.35	12.20**	6.55	13.59**	7.90 *	13.01**	11.96**	13.01**	6.86
UBS/Phillips Drew	4.86	10.11**	6.20	10.83**	0.19	10.04**	3.38	9.41 *	3.50	13.99**	3.50	8.59 *
Daiwa Europe	7.09	8.76 *	6.67	8.39 *	6.85	8.43 *	9.14 *	15.84 **	9.39 *	9.50 **	9.39 *	10.01**
Nikko Securities	4.40	7.91 *	4.49	13.98**	4.21	5.83	5.61	111.39**	2.45	5.17	2.45	5.93

Figure 1.1: Quarterly Returns on Equity and Bond

The figure reports the time series of quarterly gross returns equity and bonds. We use the Morgan Stanley Capital International value-weighted total return index in dollars as a proxy for the market portfolio, and the Merrill Lynch Global Government Bond return index in dollars for bond returns.

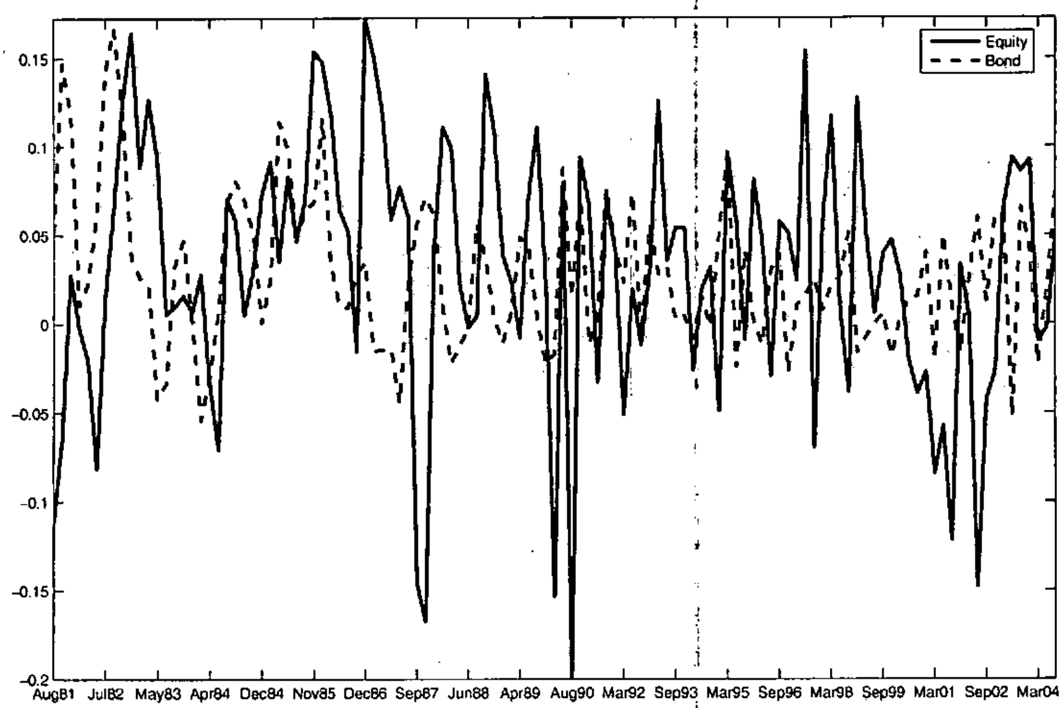


Figure 1.2: Asset Allocation

Each panel reports the time series of the optimal asset allocation recommended by each House for different asset classes.

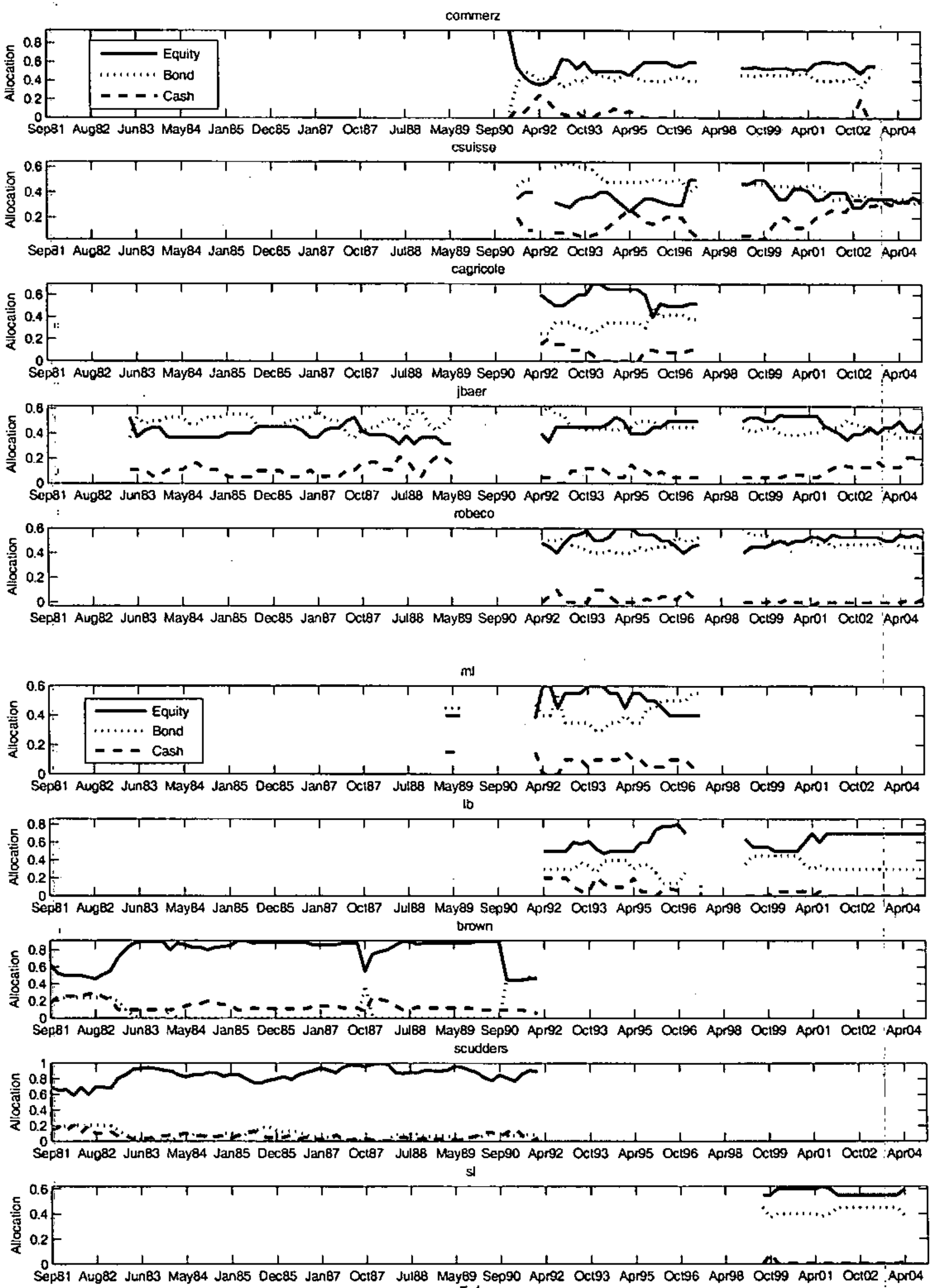


Figure 1.3: Cont'd

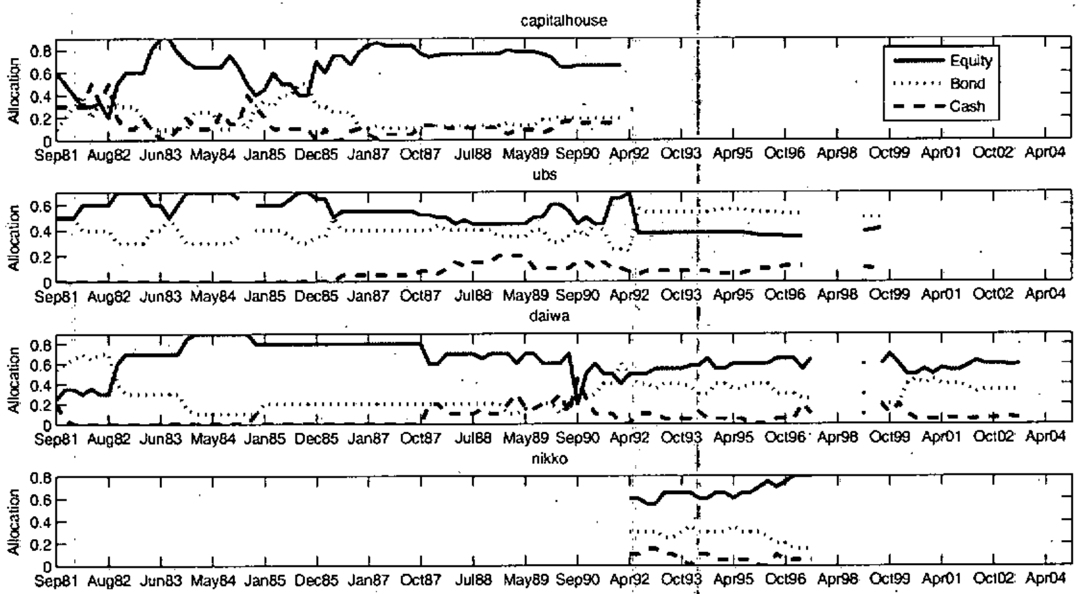
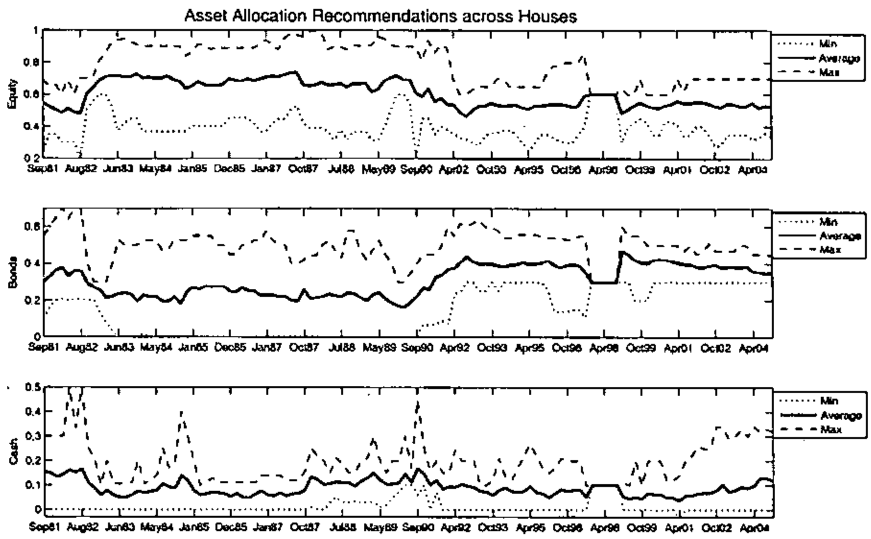


Figure 1.4: Consensus Asset Allocation

Each panel reports the time series of the minimum, average, and maximum recommended portfolio weights in equity, bond, and cash - computed across Houses.

Panel A



Panel B

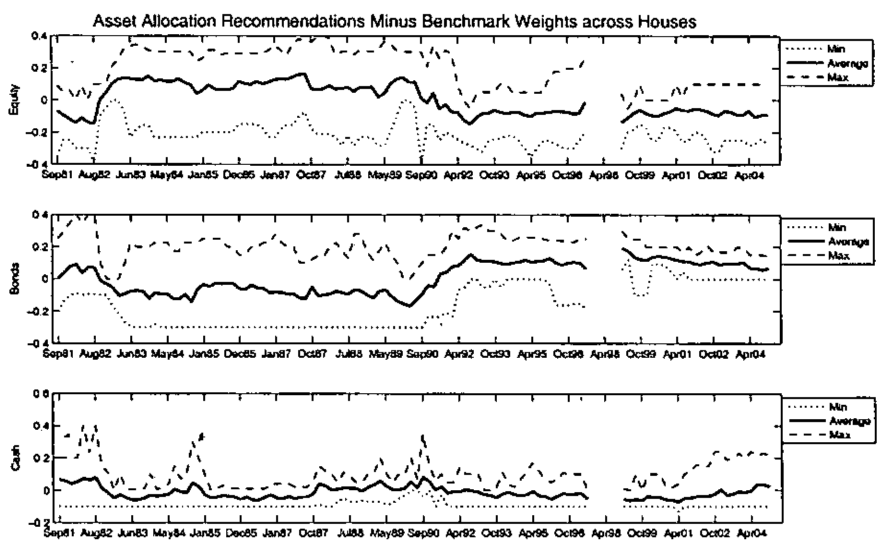


Figure 1.5: Average Portfolio Weights

Optimal portfolio composition. The upper panel plots the min, average, and max of the recommended cash/(bond+equity) allocation. The lower panel plots the min, average, and max of the recommended bond/equity allocation.

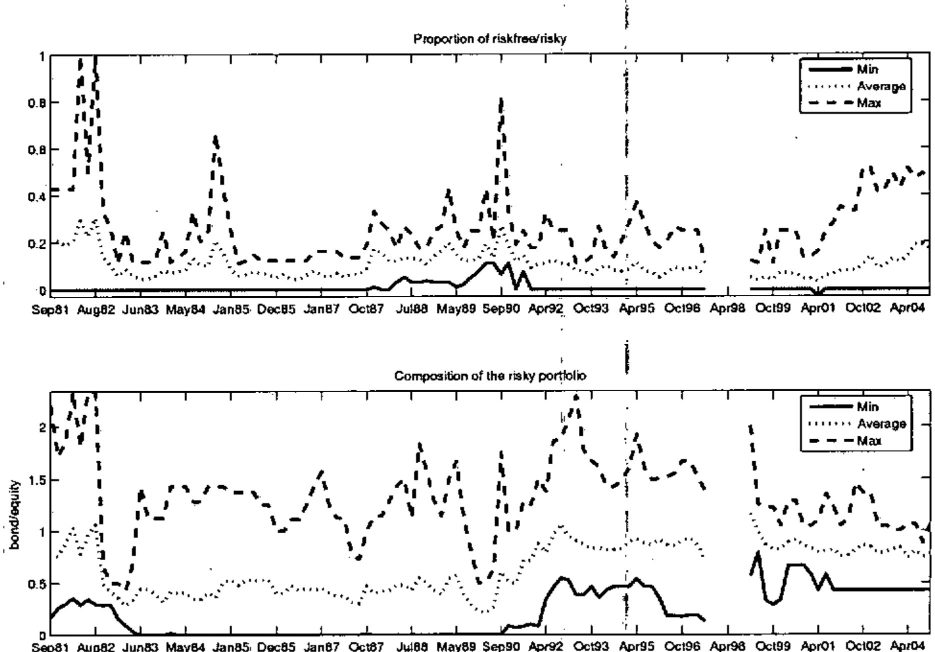
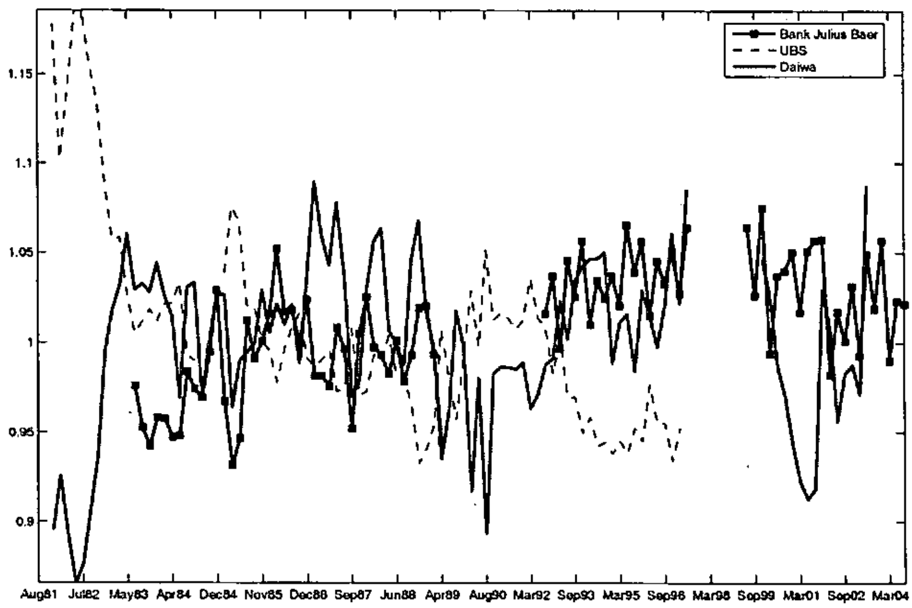


Figure 1.6: Average Beliefs about the Relative Attractiveness of Equity and Bond Markets for selected Investment Banks

The figure plots the average fitted values of s_{it+1} computed across different policy functions for selected banks. Values above 1 suggest that the equity market is more attractive compared to the bond market.



Chapter 2

Persistence in Forecasting Performance and Conditional Combination Strategies

2.1 Introduction

Forecasts[1] are of considerable importance to decision makers throughout economics and finance and are routinely used by private enterprises, government institutions and professional economists. It is therefore not surprising that much effort has gone into developing forecasting models ranging from simple, autoregressive specifications to complicated non-linear models and models with time-varying parameters. A multitude of forecasting models is typically considered because the true data generating process underlying a particular series of interest is unknown. Even the most complicated

[1]This chapter is coauthored with Allan Timmermann (UC San Diego) and it is forthcoming in the *Journal of Econometrics*. We received many helpful comments from an anonymous referee and from Eric Ghysels. We also thank seminar participants at Bocconi University and at the January 2004 San Diego conference in honor of Clive Granger.

model is likely to be misspecified and can, at best, provide a reasonable 'local' approximation to the process driving the target variable.[2]

Model instability is a source of misspecification that is likely to be particularly relevant in practice, c.f. Stock and Watson (1996). In its presence, it is highly unlikely that a single model will be dominant uniformly across time and the identity of the best local approximation is likely to change over time. If the identity of the best local model is time-varying, it is implausible that a forecasting strategy that, at each point in time, attempts to select the best current model will work well. Most obviously, if (ex-ante) the identity of the best model varies in a purely random way from period to period, it will not be possible to identify this model by considering past forecasting performance across models. Similarly, if a single best model exists but only outperforms other models by a margin that is small relative to random sampling variation, it becomes difficult to identify this model by means of statistical methods based on past performance. Even if the single best model could be identified in this situation, it is conceivable that diversification gains from combining across a set of forecasting models with similar performance will dominate the strategy of only using a single forecasting model.

In practice, the factors that give rise to long-lasting changes in the ranking of different forecasting models - e.g., major oil price shocks, policy changes, institutional shifts or market participants' learning behavior - can either take the form of discrete shocks or gradually evolving shifts and one may expect the relative performance of forecasting models to display moderate degrees of persistence. How much persistence is a question of great practical relevance. Indeed, the popular strategy of assigning equal weights to the individual forecasting models (e.g., Clemen (1989)) becomes an optimal strategy if there is no ex-ante indication of the individual models' prospective out-of-sample forecasting performance, either because the models are of a similar quality

[2]Conditions under which the true model is selected asymptotically are quite strict, c.f. White (1990) and Sin and White (1996), and asymptotic results are unlikely to provide much guidance in situations characterized by a large cross-section of forecasting models and a short time-series dimension.

or because their (relative) performance is unstable over time.

Unfortunately, little is known about persistence in forecasting performance, so the first part of our paper considers this question, establishing 'stylized facts' by studying empirically a large cross-section of economic variables and forecasting methods.[3] We find systematic evidence of persistence among both top and bottom forecasting models, but also find evidence of 'crossings' - whereby a previously good (poor) forecasting model delivers poor (good) forecasting performance out-of-sample - among linear models.

In the presence of model misspecification of unknown form and moderate degrees of persistence in the relative performance of different forecasting models, no single econometric model can be expected to outperform all others and an attractive option is to combine forecasts from several models. In their seminal paper on forecast combinations, Bates and Granger (1969) already pointed to the importance of changes in models' relative performance over time as a determinant of the scope for combining forecasts. Key questions that arise when forecast combinations are considered is how wide a set of models to include (or, similarly, how many models to exclude), whether to estimate the combination weights, use a simple combination scheme such as equal-weighting or apply shrinkage methods. The answer to such questions depends on the distribution of (relative) forecasting performance across models and the degree of persistence and is hence closely linked to the first part of our analysis. We address these issues in the second part of the paper by comparing a wide range of combination schemes that differ along these dimensions, including a new set of conditional combination strategies.

The contributions of our paper are three-fold. First, we analyze the persistence in the relative forecasting performance of a range of linear and nonlinear models using a large international data set. Second, we propose a new four-stage approach for model combination that (i) sorts models

[3]Stock and Watson (1999) consider combination methods based on expanding and rolling window estimators, two approaches that are usually associated with a stable and unstable data generating process, respectively.

into clusters based on their past performance; (ii) pools forecasts within each cluster; (iii) estimates the optimal forecast combination weights for these clusters; and (iv) shrinks the least squares combination weights towards equal weights. Third, we investigate empirically the out-of-sample forecasting performance of this new combination method and compares it with existing ones. Using a range of economic variables in the G7 countries we find that our approach improves upon existing combination methods.

The paper is organized as follows. Section 2 studies the persistence of forecasting performance across a set of linear and nonlinear time-series models. Section 3 introduces the forecast combination problem and studies the out-of-sample forecasting performance of a range of standard combination methods proposed in the literature as well as our new four-stage combination method. Section 4 concludes.

2.2 Persistence in Forecasting Performance

2.2.1 Data Set

The seven-country data set that we use is the same as that used in Stock and Watson (2004). It consists of up to 43 quarterly time series for each of the G7 economies (Canada, France, Germany, Italy, Japan, UK, and the US) over the period 1959.I – 1999.IV, although some series are available only for a shorter period. The 43 series comprise a range of asset prices (including returns, interest rates and spreads); measures of real economic activity; wages and prices; and various measures of the money stock.[4] In some cases we use more than one transformation of a given series. For example, interest rates are used both in levels and in first differences. Counting all the constructed

[4]Following Stock and Watson (2004) the variables were subject to the following transformations. First, in a few cases the series contained a large outlier–such as spikes associated with strikes–and these outliers were replaced by interpolated values. Second, series that showed significant seasonal variation were seasonally adjusted using a linear approximation to X11 in order to avoid problems with non-linearities, c.f. Ghysels, Granger and Siklos (1996). Third, data series available on a monthly basis were aggregated to get quarterly observations.

variables (such as spreads) and different transformations of the same variable, the maximum number of time series per country is 75.

2.2.2 Forecasting Models and Methods

h-step ahead forecasts of the conditional mean of the target variable, Y, are generated by time-series models of the form

$$y_{t+h} = f_i\left(\mathbf{x}_t; \theta_{i,h}\right) + \epsilon_{t+h,t,i}.$$

Here i is an index for the forecasting model, $\theta_{i,h}$ is a vector of unknown parameters, $\varepsilon_{t+h,t,i}$ is an $h-$step error term and $\mathbf{x}_t$ is a vector of predictor variables that are known at time t and may include y_t. In general, individual forecasting models only use a subset of the elements of $\mathbf{x}_t$. All forecasts are computed recursively out-of-sample, so the forecast of y_{t+h} by the ith model is computed as $f_i\left(\mathbf{x}_t; \hat{\theta}_{i,h,t}\right)$, where $\hat{\theta}_{i,h,t}$ is the estimate of $\theta_{i,h}$ given information available at time t.

Following the analysis of Stock and Watson (1999), we consider both linear and non-linear forecasting models. The class of linear models comprises simple autoregressions with lag lengths selected recursively using the Bayes information criterion (BIC), including up to four lags:

$$y_{t+h} = c + A(L)\, y_t + \epsilon_{t+h}.$$

We also consider bivariate autoregressive models that include a single additional regressor, x_t, which is an element of $\mathbf{x}_t$:

$$y_{t+h} = c + A(L)\, y_t + B(L)\, x_t + \epsilon_{t+h}.$$

Lag lengths are again selected recursively using the BIC with between 1 and 4 lags of x_t and between 0 and 4 lags of y_t. The average number of linear model specifications varies across series and countries. For example, it ranges from 36 for France, 38 for Italy, 43 for the UK, 44 for Canada and Germany, 51 for Japan to 67 for the US.

The class of non-linear forecasting models includes many of the models considered in Terasvirta, Tjostheim and Granger (1994). It includes 18 Artificial Neural Network (ANN) models with one and two hidden layers and different numbers of lags, p. Single layer feedforward neural network models take the form

$$y_{t+h} = \beta_0' \zeta_t + \sum_{i=1}^{n_1} \gamma_{1i} g\left(\beta_{1i}' \zeta_t\right) + \epsilon_{t+h},$$

$$\zeta_t = \left(1, y_t, y_{t-1}, ..., y_{t-p-1}\right), \quad p = 1, 2, 3,$$

where $g(z) = \frac{1}{1+e^z}$ is the logistic function. Neural network models with two hidden layers take the form

$$y_{t+h} = \beta_0' \zeta_t + \sum_{j=1}^{n_2} \gamma_{2j} g\left[\sum_{i=1}^{n_1} \beta_{2ji} g\left(\beta_{1i}' \zeta_t\right)\right] + \epsilon_{t+h}.$$

Our choice of design parameters for the single hidden layer ANN models are $n_1 = 1, 2, 3$ and $p = 1, 2, 3$, giving a total of nine basic models. Our choice for the ANN models with two hidden layers are $n_1 = 2, n_2 = 1, 2, 3$ and $p = 1, 2, 3$, producing nine basic models. These choices cover many of the basic neural net designs, c.f. Swanson and White (1995, 1997).[5]

[5]For all ANN models, coefficients were estimated by recursive non-linear least squares, minimizing the objective function by an ad-hoc algorithm developed by Stock and Watson.

We also consider 15 Logistic Smooth Transition Autoregression *(LSTAR)* models:

$$
\begin{aligned}
y_{t+h} &= \alpha_0'\zeta_t + d_t\beta'\zeta_t + \epsilon_{t+h} \\
d_t &= \frac{1}{1+\exp\left(\gamma_0+\gamma_1\xi_t\right)}, \\
\zeta_t &= (1, y_t, y_{t-1}, ..., y_{t-p-1}) \quad p = 1,2,3 \\
\xi_t &\in \{y_{t-1}, y_{t-2}, y_{t-3}, \Delta y_{t-1}, \Delta^2 y_{t-1}\},
\end{aligned}
$$

where the scalar ξ_t is selected from the set in (2.2.2). LSTAR models differ by the variable used to define the transition and by the lag length p, c.f. Granger and Terasvirta (1993).

Finally, we consider time-varying autoregressions *(TVARs)* whose parameters are allowed to evolve according to a multivariate random walk:

$$
\begin{aligned}
y_{t+h} &= \theta_{t,h}'\xi_t + \varepsilon_{t+h} \\
\theta_{t,h} &= \theta_{t-1,h} + \mathbf{u}_{t,h}, \\
\mathbf{u}_{t,h} &\sim iid\left(0, \lambda^2\sigma^2\mathbf{Q}\right).
\end{aligned}
$$

Here $\sigma^2\mathbf{Q}$ is the variance of $\mathbf{u}_{t,h}$. We consider seven different values of λ in the set {0.00, 0.0025, 0.005, 0.0075, 0.010, 0.015, 0.020} and up to three lags for a total of 21 TVAR models, all of which are estimated by the Kalman filter.

To avoid extreme forecasts—a problem often associated with highly non-linear models—we implement the following trimming scheme. Forecasts exceeding four recursive standard deviations of the target variable are replaced by a recursive estimate of the unconditional mean of the dependent variable computed at the time of the forecast.

2.2.3 Sorting Windows

We implement an automatic procedure to control for missing values and outliers to produce a balanced panel of forecasts. Let T_0 be the point at which the first forecast is computed and let T be the final period. For each variable and forecast horizon, h, we produce a $((T-h-T_0+1)\times N_j)$ panel of forecasts

$$\widehat{\mathbf{Y}}_{T_0+h:T}=\begin{bmatrix}\widehat{y}^{(1)}_{T,T-h} & \widehat{y}^{(2)}_{T,T-h} & \cdots & \widehat{y}^{(N_j)}_{T,T-h}\\ \widehat{y}^{(1)}_{T-1,T-h-1} & \widehat{y}^{(2)}_{T-1,T-h-1} & \cdots & \widehat{y}^{(N_j)}_{T-1,T-h-1}\\ \vdots & \vdots & \ddots & \vdots\\ \widehat{y}^{(1)}_{T_0+h,T_0} & \widehat{y}^{(2)}_{T_0+h,T_0} & \cdots & \widehat{y}^{(N_j)}_{T_0+h,T_0}\end{bmatrix},$$

where $\widehat{y}^{(i)}_{t,t+h}$ is the h-step ahead forecast computed under the ith model at time t. The superscript, i, tracks the model, $i=1,...,N_j$, and N_j is the number of models for country or forecasting method j.[6]

The $h-$period performance of the ith forecasting model at time t is measured through the loss function, $L^{(i)}_{t+h,t}\equiv L\left(y_{t+h},\widehat{y}^{(i)}_{t+h,t}\right)$. In line with common practice, we assume mean squared forecast error (MSFE) loss:

$$L(y_{t+h},\widehat{y}_{t+h,t,i})=\left(e^{(i)}_{t+h,t}\right)^2,$$

where $e^{(i)}_{t+h,t}=y_{t+h}-\widehat{y}^{(i)}_{t+h,t}$ is the $h-$step forecast error associated with the ith model's prediction at time t. We track the historical forecasting performance over a window of the last win periods by computing $S^{(i)}_t=(1/win)\sum_{\tau=t-win+1}^{t}L^{(i)}_{\tau,\tau-h}$. To study how persistent forecasting performance is

[6]For each country some series are available only for shorter subsamples so instead of dropping these series from the panel we trade off the time-series and cross-sectional dimension. Given a $T\times N$ panel we minimize the following loss function with respect to $\alpha\in(0,1)$,

$$\alpha^*=\arg\min_{\alpha}L(\alpha)=(T-T(\alpha))^2+(N-N(\alpha))^2.$$

If a time series has more than αT missing values we drop it. This gives us a $T(\alpha^*)\times N(\alpha^*)$ panel of forecasts.

through time, we consider three different tracking or 'sorting' windows used to rank the forecasting models based on their historical performance:

1. Short window: $win = 1 : S_t^{(i)} = \left(e_{t,t-h}^{(i)}\right)^2$;
2. Rolling window: $win = 20 : S_t^{(i)} = (1/win) \sum_{\tau=t-win+1}^{t} \left(e_{\tau,\tau-h}^{(i)}\right)^2$;
3. Expanding window: $S_t^{(i)} = (1/(t-h-T_0+1)) \sum_{\tau=T_0+h}^{t} \left(e_{\tau,\tau-h}^{(i)}\right)^2$.

For all methods the future out-of-sample performance at time t is based on the h-period loss, $L_{t+h,t}^{(i)}$. For each model, i, we record its rank at time t, $\mathcal{R}_{it} = f(S_t^{(1)}, ..., S_t^{(N_j)})$. The model with the best MSFE forecasting performance, gets a rank of 1, the second best a rank of 2 and so on. Using these rank orders, we sort the models into quartiles and use 4×4 contingency tables to cross-tabulate the forecasting models' sorting-period performance against their out-of-sample forecasting performance.

2.2.4 Empirical Evidence

Tables 1-4 report empirical evidence on persistence in forecasting performance for the three sorting windows using linear (Tables 1 and 2) or nonlinear (Tables 3 and 4) models and forecast horizons of $h = 1, 2, 4, 8$ quarters. Transition probability estimates, $\hat{P}_{ij}$, in these tables provide the probability of moving from quartile i (based on historical performance up to time t) to quartile j (based on future performance, $e_{t+h,t}$, $t = T_0, ..., T-h$). We show only the top corners of the tables (i.e., $\hat{P}_{11}, \hat{P}_{14}, \hat{P}_{41}, \hat{P}_{44}$) since these effectively convey information about persistence or 'anti-persistence' in forecasting performance.

Under the null of no forecasting persistence, we have $P_{ij} = 0.25$ for all i, j since the probability of good (or bad) future performance should be unaffected by past performance. Persistent forecasting performance would lead to estimates of P_{11} and P_{44} above 0.25, while P_{14} and P_{41} (the probability that a historically good model becomes a poor future model or vice versa) should be well below 0.25. Conversely, anti-persistence corresponds to small values of P_{11} and P_{44} and large values of P_{14} and P_{41}. A chi-squared test statistic can easily be constructed for the estimated transition probabilities when $h = 1$. However, at longer horizons ($h \geq 2$) the data is overlapping so the performance statistics are serially correlated. To assess the statistical significance in this situation, we therefore use a bootstrap procedure to construct confidence intervals for the transition probability estimates, $\hat{P}_{ij}$. The proportion of transition probability estimates exceeding 0.25 with a p-value below 5% is reported in Tables 2 and 4 for linear and nonlinear models, respectively.

Several interesting results emerge from the tables. First, there is robust evidence of persistence among the linear forecasting models (Table 1). Across all sorting windows, forecast horizons and countries the average estimate of P_{11} is 0.30 with 76% of the estimates exceeding 25% at a statistically significant margin. Similar numbers are obtained for the worst performing models where the average estimate of P_{44} - averaged across countries, forecast horizons and sorting windows - is 0.30 with 73% of the estimates exceeding 0.25 at the 5% significance level.

The average estimate of P_{14} is 0.29 with 75% of the estimates being significantly greater than 0.25, suggesting that there is also a high chance that the historically best performing models become the future worst models. There is clearly a smaller chance of the reverse happening - i.e., that the historically worst models become the best future models - as the average estimate of P_{41} is 0.27 and only 52% of these estimates exceed 0.25 at the 5% critical level.

There is also considerable variation in the results across sorting windows. Persistence is systematically weaker the longer the sorting window, consistent with what one would expect under model instability. Going from an expanding via a rolling to a short sorting window, the average estimate of P_{11} rises from 0.29 to 0.30 and 0.31 with 60%, 76% and 91% of these estimates being significantly greater than 0.25. A similar pattern is observed in the average estimate of P_{44} which rises from 0.28 to 0.30 and 0.33 (with 49%, 72% and 98% of these estimates being significantly greater than 0.25 at the 5% level) and in the estimates of P_{14} and P_{41} with the former rising from 0.29 to 0.31 and the latter rising from 0.26 to 0.29 as the sorting window is shortened.

Results are largely invariant with respect to the forecast horizons (h) where both $\hat{P}_{11}$ and $\hat{P}_{44}$ are close to 0.30 irrespective of the value of h.[7] Between 73% and 78% of the $\hat{P}_{11}-$estimates and between 69% and 76% of the $\hat{P}_{44}-$estimates are significant at the 5% critical level.

Disaggregating the results by country, many interesting variations are observed in our data. The mean estimate of P_{11} (averaged across series, sorting windows and forecast horizons) is 0.30 for all countries, with the estimates for Japan and the US taking the smallest values. Among the worst models, the smallest persistence, measured by the average value of $\hat{P}_{44}$, is 0.28 for the US while the largest value is 0.32, recorded for Japan. This suggests that the weakest persistence is generally found in US time series. Large variations across sorting windows are also observed. For example, for the US forecasts at the shortest horizon ($h = 1$) the proportion of estimates of P_{44} that is significant at the 5% critical level increases from 8% to 68% and 97% as we move from the expanding via the rolling to the short sorting window.

Turning to the nonlinear models (Tables 3 and 4), there is generally a lower probability of 'crossings' and fewer of the off-diagonal transition probability estimates exceed 0.25, the average value of $\hat{P}_{14}$ and $\hat{P}_{41}$ being 0.22 and 0.25, respectively. Only 15% and 35% of these estimates are significant at the 5% critical level. Overall, persistence among top models is reduced to 0.26 with only 39% (compared with 76% in the case of the linear models) of the $\hat{P}_{11}-$values significantly greater than 0.25. There is stronger persistence among the worst nonlinear models than was found for the worst linear models, however: the average estimate of P_{44} is 0.33 and 85% of the estimates

[7] To save space, disaggregated results are not reported here, but these results are available on request from the authors.

of P_{44} from the nonlinear models exceed 0.25 at the 5% critical level (compared to 73% for the linear models).

To verify the robustness of our results we also partitioned the forecasting models into groups of three based on their previous forecasting performance. We found very similar results with persistence in the top and bottom models' forecasting performance, stronger persistence among the worst linear forecasting models the shorter the sorting window and stronger persistence among the worst nonlinear forecasting models than for the linear forecasts.

We conclude the following from these findings. First, there is systematic evidence of persistence in forecasting performance both at the top end and at the bottom end of the rankings. Complicating the picture, however there is unfortunately also a strong tendency for the previous best linear models to become future underperformers. Third, in general we observe stronger persistence in the forecasting performance of the worst non-linear models and a much lower probability of crossings in these models' forecasting performance than we observed for the linear models.

2.3 Forecast Combinations

The empirical evidence reported in Section 2 suggests that there is systematic persistence in the relative forecasting performance of standard time-series models. The extent to which this evidence can be translated into improved out-of-sample forecasting performance is still an open question, however. The moderate (albeit statistically significant) degree of persistence observed among top and bottom models suggests that a strategy of using a single 'top' model is unlikely to work well and that averaging across models could improve forecasting performance. As argued earlier, persistence in forecasting performance is likely to be a key determinant of the optimal degree of averaging across models, with less averaging required the higher the degree of persistence in forecasting performance since this makes it easier to identify the best models from their historical track record.

Under mean squared error (MSE) loss the general forecast combination problem can be posed as that of choosing a mapping from a vector of N predictions $\hat{\mathbf{y}}_{t+h,t} = (\hat{y}^{(1)}_{t+h,t}, \hat{y}^{(2)}_{t+h,t}, ..., \hat{y}^{(N)}_{t+h,t})'$ to the real line, that best approximates the conditional expectation, $E[y_{t+h}|\hat{\mathbf{y}}_{t+h,t}]$. This general class of combination schemes comprises non-linear and time-varying combination methods, but it is far more common to limit the analysis by assuming a linear combination and choosing weights, $\omega_{t,h} = (\omega^{(1)}_{t,h}, ..., \omega^{(N)}_{t,h})'$ to produce a combined forecast, $\hat{y}^c_{t+h,t} = \omega'_{t,h}\hat{\mathbf{y}}_{t+h,t}$, resulting in the forecast error $e^c_{t+h,t} = y_{t+h} - \hat{y}^c_{t+h,t}$.

Assuming again that the forecaster's loss function $L(.)$ only depends on the forecast error, $e^c_{t+h,t}$, the optimal combination weights, ω^*_{th}, solve the problem

$$\omega^*_{th} = \arg\min_{\omega_{th}} E\left[L\left(e^c_{t+h,t}\right)|\hat{\mathbf{y}}_{t+h,t}\right]. \tag{2.-6}$$

Under MSE loss, $L(e) = e^2$, the combination weights are easy to characterize in population and only depend on the first two conditional moments of the joint distribution of y_{t+h} and $\hat{\mathbf{y}}_{t+h,t}$,

$$\begin{pmatrix} y_{t+h} \\ \hat{\mathbf{y}}_{t+h,t} \end{pmatrix} \sim \left(\begin{pmatrix} \mu_{yth} \\ \mu_{th} \end{pmatrix} \begin{pmatrix} \sigma^2_{yth} & \sigma'_{y\hat{y}th} \\ \sigma_{y\hat{y}th} & \Sigma_{\hat{y}\hat{y}th} \end{pmatrix} \right).$$

Assuming that $\Sigma_{\hat{y}\hat{y}th}$ is invertible, the solution to equation (2.3) is

$$\omega^*_{th} = (\mu_{th}\mu'_{th} + \Sigma^{-1}_{\hat{y}\hat{y}th})(\mu_{th}\mu_{yth} + \sigma_{y\hat{y}th}).$$

If y_{t+h} is projected on a constant as well as on the forecasts, $\hat{\mathbf{y}}_{t+h,t}$,[8] the optimal (population)

[8]The inclusion of a constant to capture bias effects is a strategy recommended (under MSE loss) by Granger and Ramanathan (1984) and, for a variety of loss functions, by Elliott and Timmermann (2004). Ruling out that the covariance matrix, $\Sigma_{\hat{y}\hat{y}th}$, is singular is innocuous here since one can always drop superfluous forecasts from the combination. One could alternatively consider non-linear combination schemes that do not impose this restriction

values of the constant and the combination weights, ω_{0th}^{c*} and ω_{th}^{*}, are

$$\omega_{0th}^{c*} = \mu_{yth} - \omega_{th}^{\prime}\mu_{th}$$

$$\omega_{th}^{*} = \Sigma_{\hat{y}\hat{y}th}^{-1}\sigma_{y\hat{y}th}.$$

These weights depend on the full conditional covariance matrix of forecasts, $\Sigma_{\hat{y}\hat{y}th}$. However, given a large number of forecasting models (N) relative to the number of time-series observations (T), it is generally not feasible or desirable to estimate optimal combination weights at the level of the individual forecasts.

A special case of (2.3) arises when one model - e.g. the ith model - has a much smaller forecast error than the other models. In this case, to an approximation, only a single forecast gets selected:

$$\omega_{th}^{*} \approx \vartheta_i,$$

where ϑ_i is an $N-$vector with zeros everywhere except for unity in the ith place.

The opposite case arises when the forecasting errors are all (roughly) of the same size with similar correlations, in which case

$$\omega_t^{*} \approx \iota_N/N,$$

where ι_N is an $N-$vector of ones. It is often found in the empirical literature that estimated "optimal" combination weights based on (2.3) lead to worse forecasting performance than such simple equal-weighted averages, (2.3), c.f. Clemen (1989).

and allow individual forecasts to be perfectly linearly correlated as long as non-linear transformations of the forecasts are not perfectly correlated.

2.3.1 Conditional Forecast Combination Strategies

Standard model selection schemes such as (2.3) and forecast combination schemes such as (2.3) or (2.3) suffer from a number of problems. With N large relative to T, estimation of the "optimal" combination weights (2.3) is either not feasible or is surrounded by considerable sampling error. While the forecasting methods in (2.3) and (2.3) do not suffer from this problem, they ignore correlation structure across different forecasts and do not efficiently use all information in the joint distribution of the forecast errors. For this reason, we propose a range of new (conditional) combination strategies that in a first stage sort the forecasting models into groups based on their recent historical forecasting performance, then pool forecasts within groups and finally combine the pooled forecasts for selected groups of models using least squares estimates of the combination weights followed by shrinkage towards equal weights.

Combination of Forecasts from Pre-selected Quartiles

The first set of combination methods operates at the level of quartile-sorted forecasts and uses information on the estimated transition probabilities to select which quartiles to include in the combination. Forecasting models are initially assigned to quartiles based on their historical forecasting performance up to the point of the prediction, t. For each quartile, a pooled (average) forecast is then computed. If the transition probability estimates (using information up to time t) suggest that a particular quartile of models produced better than average forecasts, then the pooled forecast from models in this quartile is included in the combination.

Pooling by quartile reduces the set of forecasts to a number between one and four. This is a number that is small enough to let us consider estimating optimal combination weights by least squares. We also consider shrinking the least-squares estimates of the combination weights towards equal-weights, c.f. Diebold and Pauly (1990):

$$\widehat{\omega}_t(\hat{\psi}_t) = \hat{\psi}_t\widehat{\omega}_t^{OLS} + (1-\hat{\psi}_t)\overline{\omega},$$

where $\hat{\psi}_t$, the parameter governing the amount of shrinkage, is a function of the data. This estimator shrinks the least squares estimate of the combination weights, $\widehat{\omega}_t^{OLS}$, towards equal weights, $\overline{\omega}$. As an extreme case, this includes simply using equal weights ($\hat{\psi}_t \approx 0$). Shrinkage estimators have been found to improve the finite sample performance of forecast combinations.[9]

To set out the combination strategy, let $\hat{\mathbf{y}}^q_{t+h,t}$ be the $N_q \times 1$ vector containing the forecasts belonging to quartile q, where N_q is the number of models in quartile q. We use the persistence information contained in the estimated transition probabilities at time t, $\hat{P}_{ijt}$, to select quartiles as follows:

If $\hat{P}_{11t} > \hat{P}_{14t}$: include the pooled forecast from models in the top quartile.

If $\hat{P}_{21t} + \hat{P}_{22t} > \hat{P}_{23t} + \hat{P}_{24t}$: include the pooled forecast from models in the second quartile.

If $\hat{P}_{31t} + \hat{P}_{32t} > \hat{P}_{33t} + \hat{P}_{34t}$: include the pooled forecasts from models in the third quartile.

If $\hat{P}_{41t} > \hat{P}_{44t}$: include the pooled forecasts from models in the fourth quartile.

Let $\mathcal{I}_{it}$ be an indicator variable taking the value 1 if the ith quartile is included at time t and otherwise zero, while ι_{N_q} is an $N_q \times 1$ vector of ones. Then we consider four types of combination weights applied to the forecasts pooled into quartiles, namely previous best (PB), equal-weighted (EW), optimally weighted (OW) and shrinkage-weighted (SW) combinations:

1. $PB: \hat{y}^c_{t+h,t} = (\iota'_{N_1}/N_1)\hat{\mathbf{y}}^1_{t+h,t}$.

[9] Elliott (2002) establishes conditions under which the expected loss from averaging gets closer to the expected loss from using the optimal weights as the number of forecasts (N) increases.

2. $EW : \hat{y}^c_{t+h,t} = (1/4)\sum_{q=1}^{4}(\iota'_{N_q}/N_q)\hat{\mathbf{y}}^q_{t+h,t}$.

3. $OW : \hat{y}^c_{t+h,t} = \sum_{q=1}^{4}\mathcal{I}_{qt}\hat{\omega}_{qt}\left[(\iota'_{N_q}/N_q)\hat{\mathbf{y}}^q_{t+h,t}\right]$, where $\hat{\omega}_{qt}$ are least squares estimates of the optimal combination weights for the included quartiles, computed using information up to time t.

4. $SW : \hat{y}^c_{t+h,t} = \sum_{q=1}^{4}\mathcal{I}_{qt}\hat{s}_{qt}\left[(\iota'_{N_q}/N_q)\hat{\mathbf{y}}^q_{t+h,t}\right]$, where $\hat{s}_{qt}$ are shrinkage weights applied to the selected quartiles, computed as $\hat{s}_{qt} = \psi_t\hat{\omega}_{qt} + (1-\psi_t)\left(\sum_{q=1}^{4}\mathcal{I}_{qt}\right)^{-1}$, $\psi_t = \max\left\{0, 1-\kappa\left(\frac{\sum_{q=1}^{4}\mathcal{I}_{qt}}{t-h-T_0-\sum_{q=1}^{4}\mathcal{I}_{qt}}\right)\right\}$.

Quartile-sorted combinations are referred to as $Q(W,Z)$ where $W \in \{PB, EW, OW, SW\}$ and $Z \in \{L, M, H\}$ captures the degree of shrinkage. As κ goes up, ψ_t declines and the degree of shrinkage increases so the choices of $\kappa = 2.5, 5, 7.5$ represent low, medium and strong shrinkage. These are denoted by $Q(SW,L)$, $Q(SW,M)$ and $Q(SW,H)$, respectively.[10] If none of the quartiles passes the test in the first step, we set $Q(OW) = Q(SW,.) = \frac{1}{N}\sum_{i=1}^{N}\hat{y}^{(i)}_{t+h,t}$ and average across all forecasting models.

Application of these methods requires that part of the out-of-sample period is used to establish an initial ranking of the models. We use the first 20 out-of-sample observations as our initial sorting period.

Clustering by K-mean algorithm

The approach of sorting models into quartiles can be criticized for using arbitrary cut-off points. Two models with very similar in-sample forecasting performance may get assigned to different quartiles with different weights. To deal with this problem, we propose to use a $K-$mean clustering

[10]These values are higher than the values (0.25, 0.5 and 1) considered by Stock and Watson (2004). This is because we apply shrinkage to the grouped (quartile) forecasts whereas Stock and Watson apply shrinkage to the original set of models (N) so the ratio of the number of forecasts to the effective sample size is much larger in their application than in ours. Hence we need larger values of κ to accomplish a similar degree of shrinkage.

algorithm that divides the models into a finite number of clusters based on their past forecasting performance.

To motivate this approach, Figure 1 plots the in-sample MSFE performance for output growth (up to period T_0) against the out-of-sample forecasting performance across linear forecasting models while Figure 2 does the same for the non-linear models. In general there is not much support for a simple monotonic or linear relationship between past and future forecasting performance. However, there are indications of performance clusters in some countries, notably France, Germany and Japan. There is also some evidence - notably for Italy and Japan - that the models with the very worst in-sample forecasting performance tend to generate the highest out-of-sample MSFE-values. This suggests trimming the worst models prior to computing forecasts. Trimming is particularly appealing if many models underperform the unconditional mean forecast but may also work more generally if there is a large and persistent spread in the forecasting performance across models.[11]

Suppose we identify K clusters and let $\hat{\mathbf{y}}^k_{t+h,t}$ be the $N_k \times 1$ vector containing the subset of forecasts belonging to cluster $k \in \{1, ..., K\}$ where the first cluster contains the models with the lowest historical MSFE values. We consider the following conditional combination strategies:

1. $PB : \hat{y}^c_{t+h,t} = (\iota'_{N_1}/N_1)\hat{\mathbf{y}}^1_{t+h,t}$ select the cluster with the lowest in-sample MSFE-values and use the simple mean of the forecasts in this cluster.

2. $EW : \hat{y}^c_{t+h,t} = \frac{1}{K-1}\sum_{k=1}^{K-1}(\iota'_{N_k}/N_k)\hat{\mathbf{y}}^k_{t+h,t}$ exclude the worst cluster and apply equal-weights to the forecasts from the top $K-1$ clusters.

3. $OW : \hat{y}^c_{t+h,t} = \sum_{k=1}^{K}\hat{\omega}_{kt}\left[(\iota'_{N_k}/N_k)\hat{\mathbf{y}}^k_{t+h,t}\right]$, where $\hat{\omega}_{kt}$ are least-squares estimates of the optimal combination weights for the K clusters computed using information up to period t.

4. $SW : \hat{y}^c_{t+h,t} = \sum_{k=1}^{K}\hat{s}_{kt}\left[(\iota'_{N_k}/N_k)\hat{\mathbf{y}}^k_{t+h,t}\right]$, where $\hat{s}_{kt}$ are the shrinkage weights for the

[11] See also Aiolfi and Favero (2003) and Granger and Jeon (2004) who argue in favor of trimming the worst models followed by computation of a simple equal-weighted average of the remaining forecasts.

K clusters, computed as: $\hat{s}_{kt} = \psi_t\hat{\omega}_{kt} + (1-\psi_t)\frac{1}{K}$, $\psi_t = \max\left\{0, 1-\kappa\left(\frac{K}{t-h-T_0-K}\right)\right\}$, $\kappa = 2.5, 5, 7.5$.

Cluster-sorted combinations are referred to as $C(K, W, Z)$ where K is the number of clusters, $W \in \{PB, EW, OW, SW\}$ and $Z \in \{L, M, H\}$ measures the degree of shrinkage. Hence we use the notation $C(K, SW, L)$, $C(K, SW, M)$ and $C(K, SW, H)$ for the cluster combination based on K clusters with low, medium and high shrinkage weights, respectively. We set $K = 2, 3$ and use either two or three clusters.

2.3.2 Empirical Results

Table 5 (linear models) and Table 6 (non-linear models) present results from a set of standard forecasting strategies (previous best single model (PB), equal-weighted average (EW) and top quartile ($Q(PB)$)) as well as from the four-step conditional combination strategies. These tables summarize the distribution of out-of-sample MSFE performance across countries and horizons.[12] Performance is reported relative to the out-of-sample MSFE performance of the previous best (PB) single model selected using an expanding sorting window.

First consider the results for the linear forecasting models (Table 5). Consistent with earlier studies we find that the equal-weighted combination ("mean" or EW forecast) produces good forecasts that dominate the forecasts from the previous best (PB) model. Interestingly, however, the better conditional combination strategies outperform the equal-weighted forecasts overall.

Comparing the overall forecasting performance across combination strategies, the best methods appear to be either least squares estimation of the combination weights for the selected quartiles followed by relatively strong shrinkage towards equal weights ($Q(SW, M)$ and $Q(SW, H)$)

[12] Our notation implies that PB is the forecast from the previous best single model while $Q(PB)$ is the average forecast from the quartile of previous best models and EW is the average forecast computed across all models.

or a simple average of forecasts in the top cluster ($C(2, PB)$ or $C(3, PB)$). Shrinkage towards equal weights systematically improves the forecasting performance.[13]

Turning to the results for the non-linear models shown in Table 6, the methods involving pooling within quartile-ranked forecasts followed by estimation of optimal combination weights and shrinkage towards equal weights ($Q(SW, M)$ and $Q(SW, H)$) or pooling within the top cluster of models continue to perform better on average than any of the other methods, including using the mean forecast or the average forecast from the top quartile of models.[14] Once again, the combination schemes with the strongest degree of shrinkage lead to the best overall forecasting performance.

The robustness of our results across linear and non-linear forecasting models is reassuring and suggests that two mechanisms lead to better forecasting performance. First, even though it is difficult to identify the top model among forecasting models with similar performance, it is possible to identify clusters of good and bad models. Second, and related to this point, provided that the models are pooled into groups based on their past performance, least squares estimation of the combination weights (which accounts for the correlation structure between forecasts) is a useful step. However, the estimated combination weights are surrounded by sufficiently large sampling errors that shrinkage towards equal weights generally improves on the forecasting performance.

2.4 Conclusion

This paper investigated the extent of persistence in forecasting performance across a large set of linear and nonlinear models. Much of the paper was exploratory since there is not, to our

[13]We do not report formal tests for significance of relative performance since the model choice is data driven. Hence the model under the null is sometimes nesting, while at other times does not nest, the models under the alternative. See also Stock and Watson (2004) for a discussion of this point.

[14]For the nonlinear models, in most cases only two clusters were clearly identified so we restrict the non-linear results to two clusters.

knowledge, much previous research on this question. We found significant evidence of persistence in forecasting performance. Models that were in the top and bottom quartiles when ranked by their recent historical performance have a higher than average chance of remaining in the top and bottom quartiles, respectively, in future periods. However, we also found systematic evidence of 'crossings'—whereby the previous best models become the future worst models or vice versa—among the linear forecasting models. The ranking of the worst forecasts tended to be more persistent for non-linear models than for linear models, possibly due to the fact that some of the nonlinear models are grossly misspecified—and more strongly affected by parameter estimation error—while the performance of the linear models tends to be more robust in this regard.

We next linked this evidence to the possibility of producing improved forecasts, arguing that it is likely that successful conditional combination strategies (which use information on past forecasting performance) can be designed given the persistence in forecasting performance documented in our paper. We proposed a set of new combination strategies that first sort models into either quartiles or clusters on the basis of the distribution of past forecasting performance across models, pool forecasts within each cluster and then estimate optimal combination weights and shrink these towards equal weights. This combination scheme makes use of many of the techniques proposed in the literature for improving forecast combinations such as trimming, pooling, optimal weighting and shrinkage estimation. We find evidence in our data that these conditional combination strategies lead to better overall forecasting performance than simpler strategies in common use such as using the previous best model or simply averaging across all forecasting models or a small subset of these.

Bibliography

Aiolfi, M. and C. A. Favero, 2003, Model Uncertainty, Thick Modeling and the Predictability of Stock Returns. Forthcoming in Journal of Forecasting.

Bates, J.M. and C.W.J. Granger, 1969, The Combination of Forecasts, Operations Research Quarterly 20, 451-468.

Clemen, R.T., 1989, Combining Forecasts: A Review and Annotated Bibliography, International Journal of Forecasting 5, 559-581.

Diebold, F.X. and P. Pauly, 1990, The Use of Prior Information in Forecast Combination, International Journal of Forecasting 6, 503-508.

Elliott, G., 2002, Forecast Combination with Many Forecasts, Mimeo, UCSD.

Elliott, G. and A. Timmermann, 2004, Optimal Forecast Combinations under General Loss Functions and Forecast Error Distributions, Journal of Econometrics 122, 47-79.

Ghysels, E., C.W.J. Granger and P.L. Siklos, 1996, Is Seasonal Adjustment a Linear of Nonlinear Data Filtering Process?, Journal of Business and Economic Statistics 14, 374-86.

Granger, C.W.J. and Y. Jeon, 2004, Thick Modeling, Economic Modelling 21, 323-343.

Granger, C.W.J. and R. Ramanathan, 1984, Improved Methods of Combining Forecasts. Journal of Forecasting 3, 197-204.

Granger, C.W.J. and T. Terasvirta, 1993, Modeling Nonlinear Economic Relationships (Oxford University Press, Oxford).

Sin, C.Y. and H. White, 1996, Information Criterion for Selecting Possibly Mis-specified Parametric Models, Journal of Econometrics 71, 201-225.

Stock, J.H. and M.W. Watson, 1996, Evidence on structural instability in macroeconomic time series relations, Journal of Business and Economic Statistics 14, 11-30.

Stock, J.H. and M.W. Watson, 1999, A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series, in: R.F. Engle and H. White, eds, Festschrift in Honour of Clive Granger, 1-44.

Stock, J.H. and M.W. Watson, 2004, Combination Forecasts of Output Growth in a Seven-Country Data Set, Journal of Forecasting 23, 405-430.

Swanson, N.R. and H. White, 1995, A Model Selection Approach to Assessing the Information in the Term Structure Using Linear Models and Artificial Neural Networks, Journal of Business and Economic Statistics 13, 265-75.

Swanson, N.R. and H. White, 1997, A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks, Review of Economics and Statistics 79, 540-550.

Terasvirta, T., D. Tjostheim, and C.W.J. Granger, 1994, Aspects of Modelling Nonlinear Time Series, in: R. Engle and D. McFadden, eds, Handbook of Econometrics, Vol. 4, (Elsevier: Amsterdam), 2919-60.

White, H., 1990, A Consistent Model Selection, in: C.W.J. Granger, ed., Modeling Economic Series, (Oxford University Press, Oxford)

Table 2.1: Transition probabilities estimated for the linear models

Each cell reports the corner probabilities of the 4x4 contingency table tracking the forecasting models' initial and subsequent h-period rankings. Transition probabilities P_{ij} give the probability of moving from quartile i (based on historical performance, $e^2_{i,t-h}$) to quartile j (based on future performance, $e^2_{i+h,t}$). All estimates are averaged across variables within a particular country.

	h=1		h=2		h=4		h=8	
Expanding Sorting Window								
USA	0.30	0.31	0.29	0.30	0.29	0.29	0.29	0.30
	0.23	0.23	0.24	0.25	0.26	0.26	0.26	0.26
UK	0.29	0.29	0.28	0.28	0.29	0.28	0.29	0.31
	0.27	0.29	0.27	0.30	0.27	0.30	0.28	0.30
France	0.28	0.28	0.28	0.27	0.28	0.28	0.28	0.28
	0.26	0.28	0.27	0.29	0.27	0.29	0.29	0.29
Germany	0.29	0.30	0.29	0.29	0.28	0.29	0.29	0.30
	0.26	0.26	0.26	0.27	0.27	0.28	0.27	0.29
Japan	0.29	0.27	0.28	0.26	0.29	0.26	0.27	0.27
	0.26	0.29	0.26	0.30	0.26	0.32	0.28	0.30
Canada	0.30	0.31	0.30	0.31	0.30	0.30	0.28	0.28
	0.24	0.25	0.24	0.25	0.26	0.27	0.27	0.27
Italy	0.28	0.28	0.28	0.27	0.27	0.27	0.29	0.28
	0.27	0.28	0.26	0.28	0.27	0.28	0.28	0.29
Rolling Sorting Window								
USA	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.31
	0.25	0.27	0.26	0.29	0.27	0.28	0.27	0.28
UK	0.29	0.29	0.30	0.29	0.30	0.30	0.30	0.31
	0.28	0.31	0.28	0.32	0.28	0.31	0.28	0.31
France	0.30	0.28	0.30	0.28	0.31	0.29	0.31	0.31
	0.28	0.31	0.27	0.31	0.27	0.31	0.28	0.30
Germany	0.31	0.30	0.30	0.31	0.30	0.32	0.31	0.32
	0.27	0.29	0.27	0.29	0.28	0.29	0.28	0.29
Japan	0.30	0.27	0.29	0.27	0.30	0.27	0.30	0.30
	0.27	0.31	0.27	0.32	0.26	0.33	0.29	0.31
Canada	0.31	0.31	0.31	0.32	0.31	0.31	0.31	0.32
	0.26	0.27	0.26	0.28	0.26	0.29	0.27	0.28
Italy	0.29	0.28	0.30	0.29	0.30	0.29	0.31	0.30
	0.28	0.31	0.27	0.32	0.28	0.31	0.28	0.31
Short Sorting Window								
USA	0.30	0.30	0.31	0.31	0.31	0.30	0.30	0.30
	0.30	0.33	0.29	0.33	0.29	0.32	0.28	0.30
UK	0.30	0.32	0.31	0.32	0.32	0.30	0.32	0.30
	0.31	0.34	0.30	0.35	0.28	0.36	0.29	0.33
France	0.32	0.30	0.33	0.30	0.32	0.30	0.34	0.30
	0.29	0.35	0.28	0.35	0.29	0.33	0.27	0.34
Germany	0.32	0.31	0.31	0.32	0.32	0.31	0.31	0.31
	0.30	0.34	0.30	0.33	0.30	0.33	0.29	0.31
Japan	0.30	0.31	0.31	0.30	0.32	0.29	0.32	0.30
	0.30	0.34	0.29	0.34	0.27	0.36	0.28	0.34
Canada	0.30	0.32	0.31	0.32	0.31	0.31	0.31	0.31
	0.30	0.32	0.29	0.32	0.29	0.32	0.28	0.31
Italy	0.32	0.30	0.32	0.30	0.33	0.28	0.33	0.30
	0.30	0.35	0.29	0.35	0.27	0.36	0.28	0.35

Table 2.2: Significance of transition probabilities estimated for the linear models

Each cell reports the percentage of corner probabilities in Table 1 that is greater than 0.25 at the 5% significance level.

	h=1		h=2		h=4		h=8	
Expanding Sorting Window								
USA	0.88	0.89	0.72	0.80	0.66	0.78	0.68	0.81
	0.05	0.08	0.21	0.23	0.32	0.36	0.30	0.42
UK	0.68	0.66	0.54	0.61	0.56	0.65	0.65	0.69
	0.46	0.52	0.39	0.66	0.44	0.67	0.53	0.67
France	0.57	0.57	0.60	0.45	0.47	0.49	0.36	0.45
	0.32	0.49	0.36	0.49	0.47	0.60	0.55	0.64
Germany	0.62	0.66	0.62	0.66	0.53	0.67	0.52	0.70
	0.30	0.40	0.36	0.40	0.57	0.63	0.58	0.64
Japan	0.65	0.47	0.50	0.39	0.55	0.35	0.48	0.53
	0.31	0.60	0.34	0.69	0.33	0.78	0.52	0.64
Canada	0.75	0.82	0.67	0.82	0.65	0.67	0.46	0.52
	0.18	0.18	0.14	0.18	0.31	0.45	0.43	0.44
Italy	0.43	0.61	0.53	0.53	0.43	0.43	0.49	0.53
	0.41	0.47	0.41	0.57	0.43	0.51	0.49	0.62
Rolling Sorting Window								
USA	0.95	0.92	0.93	0.89	0.93	0.93	0.77	0.90
	0.23	0.68	0.36	0.76	0.53	0.70	0.45	0.62
UK	0.68	0.57	0.70	0.66	0.67	0.76	0.71	0.90
	0.57	0.79	0.57	0.82	0.56	0.76	0.47	0.78
France	0.72	0.49	0.70	0.62	0.71	0.69	0.64	0.80
	0.51	0.85	0.45	0.81	0.47	0.78	0.45	0.70
Germany	0.79	0.79	0.74	0.92	0.82	0.88	0.72	0.88
	0.51	0.74	0.53	0.72	0.71	0.65	0.46	0.64
Japan	0.71	0.50	0.61	0.44	0.70	0.48	0.66	0.79
	0.40	0.81	0.50	0.85	0.47	0.85	0.69	0.76
Canada	0.88	0.98	0.93	0.96	0.84	0.93	0.85	0.93
	0.25	0.42	0.32	0.44	0.29	0.73	0.37	0.57
Italy	0.69	0.71	0.69	0.59	0.68	0.60	0.70	0.70
	0.47	0.71	0.41	0.84	0.43	0.74	0.55	0.74
Short Sorting Window								
USA	0.84	0.83	0.99	0.93	0.99	0.93	0.93	0.97
	0.83	0.97	0.81	1.00	0.84	1.00	0.81	0.92
UK	0.88	0.93	0.84	0.91	0.87	0.69	0.84	0.73
	0.93	0.98	0.80	0.96	0.59	1.00	0.61	0.96
France	0.91	0.85	0.96	0.74	0.82	0.80	0.93	0.82
	0.66	1.00	0.64	1.00	0.60	0.96	0.45	0.93
Germany	0.98	0.94	0.92	0.94	0.98	0.94	0.92	0.98
	0.89	1.00	0.85	0.96	0.84	0.98	0.70	0.90
Japan	0.73	0.84	0.87	0.69	0.90	0.62	0.91	0.81
	0.82	1.00	0.68	0.98	0.55	1.00	0.62	0.97
Canada	0.88	0.91	0.96	1.00	0.95	0.89	0.96	0.96
	0.88	0.96	0.89	1.00	0.80	0.96	0.65	0.96
Italy	0.92	0.78	0.92	0.88	0.89	0.68	0.89	0.77
	0.78	1.00	0.71	1.00	0.40	1.00	0.57	1.00

Table 2.3: Transition probabilities estimated for the non linear models

Each cell reports the corner probabilities of the 4x4 contingency table tracking the forecasting models' initial and subsequent h-period rankings. Transition probabilities P_{ij} give the probability of moving from quartile i (based on historical performance, $e^2_{t,t-h}$) to quartile j (based on future performance, $e^2_{t+h,t}$). All estimates are averaged across variables within a particular country.

	h=1		h=2		h=4		h=8	
Expanding Sorting Window								
USA	0.25	0.20	0.26	0.19	0.26	0.22	0.26	0.23
	0.25	0.34	0.25	0.33	0.26	0.31	0.25	0.30
UK	0.25	0.19	0.25	0.19	0.26	0.21	0.25	0.21
	0.25	0.35	0.25	0.33	0.26	0.33	0.25	0.33
France	0.25	0.19	0.25	0.21	0.25	0.22	0.26	0.23
	0.25	0.34	0.25	0.31	0.26	0.31	0.26	0.29
Germany	0.26	0.19	0.25	0.19	0.26	0.19	0.27	0.20
	0.24	0.35	0.25	0.33	0.24	0.34	0.23	0.35
Japan	0.25	0.18	0.25	0.19	0.26	0.20	0.27	0.23
	0.25	0.35	0.26	0.32	0.25	0.32	0.25	0.30
Canada	0.25	0.20	0.26	0.20	0.26	0.22	0.26	0.21
	0.25	0.35	0.25	0.33	0.25	0.32	0.25	0.33
Italy	0.25	0.20	0.25	0.20	0.26	0.22	0.25	0.22
	0.25	0.33	0.25	0.32	0.25	0.30	0.25	0.29
Rolling Sorting Window								
USA	0.26	0.20	0.26	0.20	0.26	0.21	0.26	0.22
	0.26	0.36	0.26	0.34	0.26	0.32	0.26	0.31
UK	0.26	0.19	0.26	0.20	0.27	0.20	0.27	0.20
	0.25	0.36	0.25	0.34	0.25	0.34	0.25	0.33
France	0.26	0.20	0.26	0.21	0.26	0.22	0.27	0.23
	0.25	0.35	0.25	0.33	0.26	0.32	0.26	0.30
Germany	0.26	0.19	0.26	0.20	0.27	0.20	0.28	0.20
	0.24	0.35	0.25	0.33	0.24	0.34	0.23	0.34
Japan	0.27	0.18	0.26	0.19	0.27	0.20	0.28	0.22
	0.25	0.36	0.25	0.34	0.25	0.33	0.25	0.31
Canada	0.25	0.20	0.26	0.21	0.26	0.22	0.26	0.22
	0.25	0.35	0.26	0.33	0.25	0.33	0.25	0.33
Italy	0.26	0.20	0.26	0.20	0.27	0.21	0.27	0.22
	0.26	0.34	0.26	0.33	0.24	0.32	0.25	0.31
Short Sorting Window								
USA	0.27	0.25	0.27	0.24	0.27	0.25	0.27	0.25
	0.26	0.34	0.26	0.33	0.27	0.31	0.25	0.30
UK	0.27	0.25	0.27	0.24	0.27	0.24	0.27	0.24
	0.26	0.35	0.26	0.34	0.25	0.32	0.26	0.32
France	0.28	0.24	0.27	0.24	0.28	0.24	0.27	0.26
	0.26	0.35	0.26	0.33	0.26	0.31	0.26	0.30
Germany	0.27	0.23	0.26	0.24	0.27	0.23	0.27	0.24
	0.26	0.35	0.25	0.33	0.25	0.32	0.25	0.30
Japan	0.27	0.24	0.28	0.24	0.28	0.24	0.29	0.24
	0.26	0.35	0.25	0.34	0.25	0.32	0.24	0.31
Canada	0.26	0.25	0.26	0.24	0.26	0.25	0.27	0.25
	0.26	0.34	0.26	0.33	0.26	0.31	0.25	0.32
Italy	0.27	0.24	0.27	0.24	0.28	0.23	0.27	0.24
	0.26	0.35	0.26	0.33	0.25	0.33	0.26	0.31

Table 2.4: Significance of transition probabilities estimated for the non linear models

Each cell reports the percentage of corner probabilities in Table 3 that is greater than 0.25 at the 5% significance level.

	h=1		h=2		h=4		h=8	
Expanding Sorting Window								
USA	0.35	0.13	0.36	0.08	0.48	0.22	0.36	0.29
	0.43	0.88	0.40	0.88	0.40	0.75	0.36	0.66
UK	0.21	0.03	0.26	0.02	0.41	0.09	0.38	0.20
	0.36	0.93	0.36	0.91	0.48	0.77	0.42	0.76
France	0.22	0.02	0.18	0.04	0.30	0.11	0.22	0.22
	0.24	0.90	0.24	0.78	0.34	0.83	0.33	0.62
Germany	0.40	0.07	0.33	0.11	0.47	0.07	0.46	0.22
	0.26	0.93	0.32	0.84	0.25	0.80	0.24	0.74
Japan	0.25	0.05	0.30	0.05	0.30	0.07	0.39	0.25
	0.33	0.94	0.44	0.87	0.30	0.77	0.36	0.55
Canada	0.32	0.14	0.32	0.08	0.35	0.19	0.34	0.21
	0.32	0.93	0.29	0.83	0.30	0.79	0.38	0.68
Italy	0.22	0.06	0.34	0.09	0.43	0.20	0.30	0.30
	0.43	0.94	0.43	0.86	0.32	0.66	0.30	0.63
Rolling Sorting Window								
USA	0.31	0.01	0.31	0.11	0.37	0.12	0.40	0.26
	0.44	0.99	0.45	0.95	0.49	0.88	0.42	0.66
UK	0.36	0.02	0.34	0.00	0.46	0.05	0.48	0.12
	0.41	0.98	0.36	0.91	0.36	0.88	0.40	0.74
France	0.31	0.00	0.31	0.06	0.28	0.04	0.38	0.22
	0.29	1.00	0.31	0.88	0.34	0.83	0.47	0.64
Germany	0.37	0.00	0.32	0.07	0.44	0.07	0.57	0.11
	0.30	0.96	0.32	0.86	0.27	0.84	0.28	0.81
Japan	0.40	0.02	0.37	0.06	0.41	0.07	0.45	0.21
	0.25	0.97	0.35	0.89	0.36	0.82	0.41	0.73
Canada	0.25	0.07	0.34	0.07	0.26	0.07	0.46	0.14
	0.32	0.95	0.39	0.95	0.35	0.86	0.39	0.68
Italy	0.41	0.04	0.50	0.02	0.50	0.14	0.53	0.20
	0.41	0.98	0.36	0.98	0.34	0.86	0.28	0.85
Short Sorting Window								
USA	0.41	0.28	0.45	0.23	0.48	0.32	0.42	0.26
	0.41	0.92	0.37	0.95	0.44	0.86	0.36	0.67
UK	0.36	0.26	0.48	0.21	0.48	0.20	0.46	0.32
	0.45	0.95	0.28	0.93	0.27	0.84	0.38	0.80
France	0.55	0.10	0.43	0.16	0.47	0.23	0.40	0.36
	0.35	0.96	0.35	0.90	0.30	0.79	0.33	0.67
Germany	0.47	0.12	0.42	0.23	0.40	0.20	0.43	0.26
	0.30	0.95	0.30	0.93	0.35	0.95	0.22	0.74
Japan	0.32	0.27	0.37	0.29	0.44	0.20	0.59	0.16
	0.40	0.95	0.30	0.90	0.25	0.89	0.21	0.75
Canada	0.34	0.20	0.39	0.20	0.42	0.25	0.46	0.23
	0.34	0.93	0.34	0.95	0.40	0.75	0.21	0.79
Italy	0.41	0.22	0.45	0.16	0.52	0.20	0.35	0.30
	0.41	0.98	0.41	0.98	0.27	0.91	0.38	0.78

Table 2.5:

Out-of-sample forecasting performance of combination schemes applied to linear models. Each panel reports the distribution of out-of-sample MSFE - relative to that of the previous best model using an expanding window - averaged across variables, countries and forecast horizons (1,560 forecasts) for different combination strategies. Standard combination strategies include PB, the previous best model, the average across the models in the top quartile $Q(PB)$, and across all models EW. Quartile and cluster-sorted conditional combination strategies are referred to as $Q(W, Z)$, and $C(K, W, Z)$ respectively, where $W \in \{EW, OW, PB, SW\}$ (equal weighted, optimally weighted, previous best, and shrinkage weighted), $Z \in \{L, M, H\}$ is the degree of shrinkage (low, medium, high), and K is the number of clusters.

Expanding Sorting Window

	Min	10%	25%	Median	75%	90%	Max	Mean
PB	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
EW	0.204	0.755	0.860	0.940	0.996	1.060	2.694	0.928
$Q(PB)$	0.223	0.762	0.866	0.939	0.991	1.036	2.360	0.923
$Q(OW)$	0.152	0.705	0.820	0.919	0.994	1.073	3.518	0.912
$Q(SW, L)$	0.251	0.701	0.821	0.917	0.985	1.056	2.820	0.904
$Q(SW, M)$	0.304	0.705	0.828	0.916	0.981	1.047	2.695	0.900
$Q(SW, H)$	0.328	0.709	0.834	0.916	0.979	1.040	2.688	0.901
$C(3, PB)$	0.145	0.762	0.870	0.944	0.993	1.038	1.951	0.924
$C(3, EW)$	0.204	0.755	0.860	0.940	0.996	1.060	2.694	0.928
$C(3, OW)$	0.271	0.808	0.934	1.045	1.240	1.552	13.748	1.187
$C(3, SW, L)$	0.292	0.780	0.897	1.000	1.120	1.313	9.226	1.058
$C(3, SW, M)$	0.203	0.765	0.872	0.969	1.051	1.171	5.966	0.980
$C(3, SW, H)$	0.211	0.760	0.866	0.953	1.019	1.097	3.818	0.945
$C(2, PB)$	0.173	0.756	0.862	0.941	0.994	1.046	2.490	0.925
$C(2, OW)$	0.198	0.741	0.864	0.976	1.110	1.389	8.287	1.057
$C(2, SW, L)$	0.237	0.734	0.855	0.958	1.070	1.264	6.009	1.006
$C(2, SW, M)$	0.270	0.731	0.851	0.947	1.037	1.182	4.228	0.968
$C(2, SW, H)$	0.213	0.737	0.847	0.938	1.020	1.129	2.943	0.943

Rolling Sorting Window

	Min	10%	25%	Median	75%	90%	Max	Mean
PB	0.510	0.874	0.948	1.005	1.060	1.133	1.967	1.010
EW	0.204	0.755	0.860	0.940	0.996	1.060	2.694	0.928
$Q(PB)$	0.163	0.754	0.863	0.940	0.995	1.049	2.408	0.924
$Q(OW)$	0.152	0.697	0.822	0.919	0.992	1.079	3.541	0.911
$Q(SW, L)$	0.253	0.694	0.822	0.916	0.986	1.055	2.824	0.903
$Q(SW, M)$	0.304	0.697	0.827	0.914	0.983	1.047	2.776	0.900
$Q(SW, H)$	0.328	0.704	0.829	0.915	0.981	1.042	2.760	0.900
$C(3, PB)$	0.172	0.757	0.864	0.942	0.995	1.048	1.922	0.926
$C(3, EW)$	0.204	0.755	0.860	0.940	0.996	1.060	2.694	0.928
$C(3, OW)$	0.222	0.845	0.971	1.114	1.331	1.662	16.616	1.263
$C(3, SW, L)$	0.235	0.802	0.924	1.035	1.186	1.389	11.452	1.113
$C(3, SW, M)$	0.272	0.778	0.892	0.992	1.093	1.229	7.642	1.017
$C(3, SW, H)$	0.304	0.763	0.877	0.969	1.044	1.149	5.023	0.970
$C(2, PB)$	0.178	0.752	0.860	0.939	0.995	1.044	2.259	0.923
$C(2, OW)$	0.245	0.763	0.895	1.019	1.189	1.464	10.246	1.131
$C(2, SW, L)$	0.251	0.747	0.882	0.994	1.135	1.332	8.357	1.064
$C(2, SW, M)$	0.283	0.742	0.872	0.977	1.091	1.238	6.710	1.013
$C(2, SW, H)$	0.295	0.740	0.863	0.963	1.058	1.171	5.303	0.977

Short Sorting Window

	Min	10%	25%	Median	75%	90%	Max	Mean
PB	0.291	0.825	0.924	1.016	1.098	1.207	3.034	1.023
EW	0.204	0.755	0.860	0.940	0.996	1.060	2.694	0.928
$Q(PB)$	0.160	0.755	0.865	0.942	0.999	1.062	2.585	0.929
$Q(OW)$	0.269	0.695	0.821	0.921	1.004	1.098	3.756	0.920
$Q(SW, L)$	0.280	0.697	0.819	0.920	0.997	1.078	3.014	0.910
$Q(SW, M)$	0.296	0.701	0.820	0.917	0.990	1.063	2.575	0.905
$Q(SW, H)$	0.281	0.708	0.827	0.917	0.985	1.054	2.590	0.904
$C(3, PB)$	0.175	0.749	0.855	0.938	0.998	1.063	2.449	0.924
$C(3, EW)$	0.204	0.755	0.860	0.940	0.996	1.060	2.694	0.928
$C(3, OW)$	0.313	0.868	1.005	1.172	1.464	2.022	15.199	1.404
$C(3, SW, L)$	0.330	0.819	0.948	1.068	1.253	1.570	9.402	1.190
$C(3, SW, M)$	0.275	0.783	0.902	1.008	1.127	1.317	5.183	1.055
$C(3, SW, H)$	0.282	0.768	0.882	0.977	1.064	1.194	3.074	0.989
$C(2, PB)$	0.189	0.749	0.856	0.934	0.992	1.058	2.625	0.923
$C(2, OW)$	0.254	0.797	0.921	1.055	1.262	1.671	9.782	1.223
$C(2, SW, L)$	0.258	0.783	0.900	1.021	1.180	1.479	7.376	1.127
$C(2, SW, M)$	0.275	0.766	0.881	0.989	1.116	1.336	5.425	1.053
$C(2, SW, H)$	0.321	0.755	0.865	0.968	1.079	1.233	3.929	1.002

Table 2.6:

Out-of-sample forecasting performance of combination schemes applied to non linear models. Each panel reports the distribution of out-of-sample MSFE - relative to that of the previous best model using an expanding window - averaged across variables, countries and forecast horizons (1,560 forecasts) for different combination strategies. Standard combination strategies include PB, the previous best model, the average across the models in the top quartile $Q(PB)$, and across all models EW. Quartile and cluster-sorted conditional combination strategies are referred to as $Q(W,Z)$, and $C(K,W,Z)$ respectively, where $W \in \{EW, OW, PB, SW\}$ (equal weighted, optimally weighted, previous best, and shrinkage weighted), $Z \in \{L, M, H\}$ is the degree of shrinkage (low, medium, high), and K is the number of clusters.

	Min	10%	25%	Median	75%	90%	Max	Mean
				Expanding Sorting Window				
PB	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
EW	0.355	0.708	0.814	0.906	0.988	1.065	1.651	0.900
$Q(PB)$	0.369	0.748	0.833	0.907	0.970	1.018	1.630	0.896
$Q(OW)$	0.089	0.641	0.746	0.858	0.968	1.062	2.925	0.862
$Q(SW,L)$	0.153	0.647	0.747	0.855	0.960	1.036	2.259	0.853
$Q(SW,M)$	0.207	0.652	0.747	0.854	0.951	1.024	1.727	0.849
$Q(SW,H)$	0.232	0.661	0.755	0.857	0.949	1.016	1.555	0.850
$C(2,PB)$	0.356	0.709	0.813	0.893	0.963	1.018	1.339	0.881
$C(2,OW)$	0.154	0.679	0.807	0.938	1.101	1.438	7.130	1.026
$C(2,SW,L)$	0.164	0.671	0.798	0.928	1.069	1.339	5.659	0.990
$C(2,SW,M)$	0.167	0.672	0.798	0.916	1.041	1.257	4.423	0.962
$C(2,SW,H)$	0.183	0.676	0.797	0.913	1.026	1.194	3.789	0.944
				Rolling Sorting Window				
PB	0.520	0.870	0.948	1.017	1.097	1.187	3.710	1.034
EW	0.355	0.708	0.814	0.906	0.988	1.065	1.651	0.900
$Q(PB)$	0.381	0.749	0.830	0.912	0.982	1.041	1.825	0.904
$Q(OW)$	0.089	0.639	0.752	0.869	0.977	1.073	2.894	0.867
$Q(SW,L)$	0.135	0.643	0.750	0.865	0.963	1.049	2.246	0.858
$Q(SW,M)$	0.143	0.649	0.752	0.860	0.957	1.033	1.808	0.853
$Q(SW,H)$	0.181	0.660	0.757	0.864	0.954	1.023	1.633	0.853
$C(2,PB)$	0.357	0.714	0.810	0.897	0.970	1.029	1.467	0.885
$C(2,OW)$	0.176	0.687	0.813	0.944	1.117	1.448	5.869	1.041
$C(2,SW,L)$	0.158	0.679	0.807	0.932	1.076	1.359	4.906	1.001
$C(2,SW,M)$	0.157	0.682	0.803	0.919	1.053	1.285	4.320	0.970
$C(2,SW,H)$	0.170	0.685	0.802	0.914	1.034	1.218	3.787	0.949
				Short Sorting Window				
PB	0.392	0.850	0.960	1.091	1.242	1.461	3.869	1.130
EW	0.355	0.708	0.814	0.906	0.988	1.065	1.651	0.900
$Q(PB)$	0.385	0.751	0.844	0.943	1.047	1.179	2.221	0.958
$Q(OW)$	0.088	0.642	0.758	0.880	0.996	1.118	2.682	0.886
$Q(SW,L)$	0.144	0.648	0.759	0.874	0.983	1.086	2.245	0.876
$Q(SW,M)$	0.166	0.649	0.762	0.874	0.976	1.070	1.880	0.871
$Q(SW,H)$	0.227	0.659	0.768	0.878	0.970	1.055	1.621	0.871
$C(2,PB)$	0.371	0.722	0.819	0.915	0.991	1.073	1.765	0.907
$C(2,OW)$	0.158	0.684	0.813	0.939	1.093	1.373	4.455	1.003
$C(2,SW,L)$	0.150	0.682	0.803	0.926	1.059	1.282	3.947	0.968
$C(2,SW,M)$	0.157	0.678	0.799	0.916	1.037	1.213	3.483	0.944
$C(2,SW,H)$	0.180	0.677	0.801	0.911	1.022	1.170	3.060	0.930

Figure 2.1:

Scatter plot of the average in-sample versus out-of-sample MSFE values generated by linear forecasting models estimated for output growth.

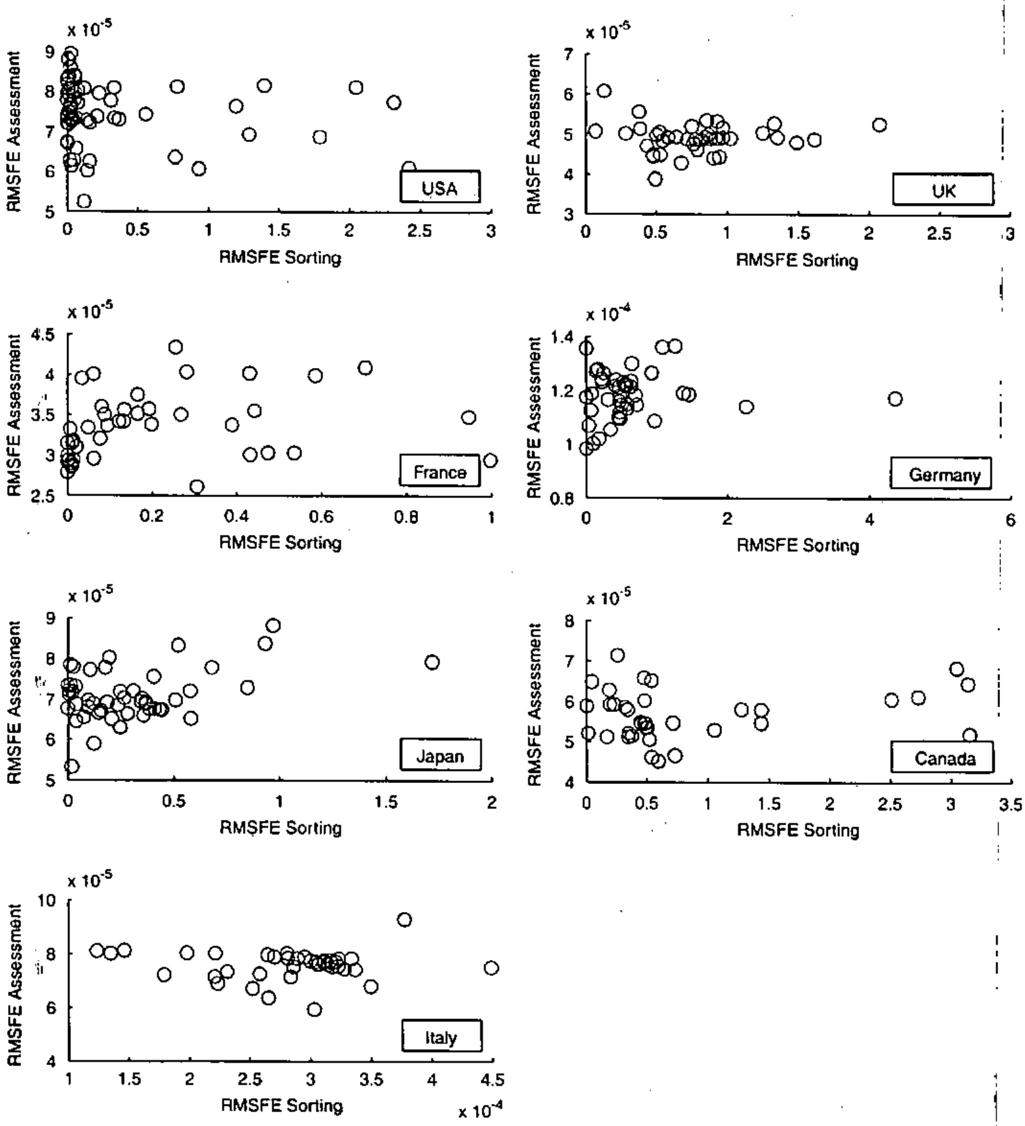
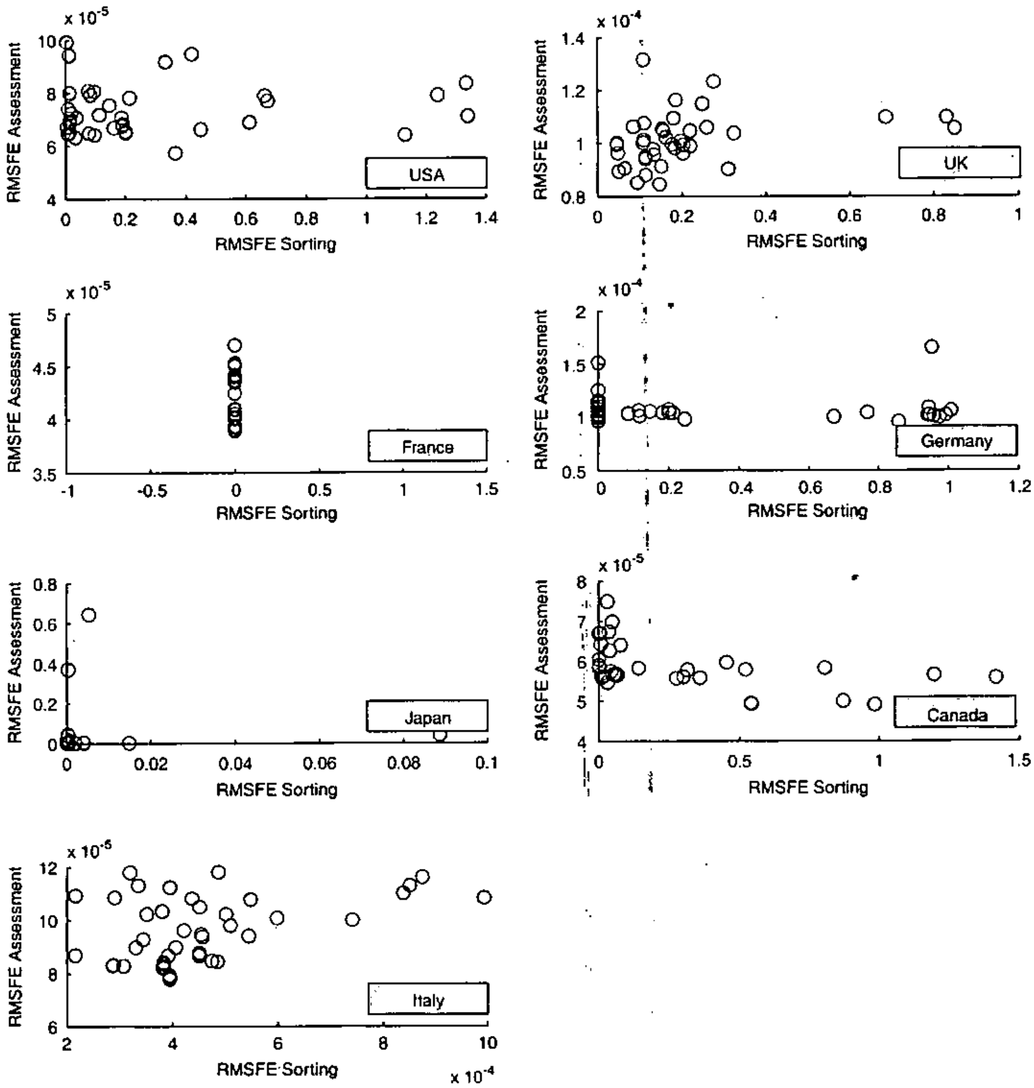


Figure 2.2:

Scatter plot of the average in-sample versus out-of-sample MSFE values generated by non-linear forecasting models estimated for output growth.



Chapter 3

Model Uncertainty, Thick Modelling and the Predictability of Stock Returns

3.1 Introduction

Recent[1] financial research has provided ample evidence on the predictability of stock returns identifying a large number of financial and macro variables that appear to predict future stock returns[2]. Even though financial economists and practitioners have agreed upon a restricted set of explanatory variables that could be used to forecast future stock returns, there is no agreement

[1]This chapter is coauthored with Carlo Favero (IGIER, Bocconi University) and it is published in the *Journal of Forecasting*, 24, 233254 (2005). We are indebted to the editor, Allan Timmermann, two anonymous referees, Francesco Corielli, Clive Granger, Wessel Marquering, Alessandro Penati, Franco Peracchi, Hashem Pesaran, Eduardo Rossi, Guido Tabellini, as well as seminar participants at 'Ente Einaudi for Monetary and Financial Studies' in Rome, and University of California, San Diego for comments and suggestions.

[2]See for example Ait-Sahalia and Brandt (2001), Avramov (2002), Bossaert and Hillion (1999), Brandt (1999), Campbell and Shiller (1988a, 1988b), Cochrane (1999), Fama and French (1988), Keim and Stambaugh (1986), Lamont (1998), Lander et al.(1997), Lettau and Ludvigson (2001), Pesaran and Timmermann (1995, 2002).

on the use of a single specification. Different attempts have been made to come up with a robust specification.

Pesaran and Timmermann (1995) (henceforth, P&T) consider a time-varying parameterization for the forecasting model to find that the predictive power of various economic factors over stock returns changes through time and tends to vary with the volatility of returns. They apply a 'recursive modelling' approach, according to which at each point in time all the possible forecasting models are estimated and returns are predicted by relying on the best model, chosen on the basis of some given in-sample statistical criterion. The dynamic portfolio allocation, based on the signal generated by a time-varying model for asset returns, is shown to out-perform the buy-and-hold strategy over the period 1959-1992. The results obtained for the US are successfully replicated in a recent paper concentrating on the UK evidence, Pesaran and Timmermann (2000). Following this line of research Bossaerts and Hillion (1999) implement different model selection criteria in order to verify the evidence of the predictability in excess returns, discovering that even the best prediction models has no out-of -sample predicting power.

The standard practice of choosing the best specification according to some selection criterion can be labelled as thin modeling because a single forecast is associated with all available specifications. In reality a generic investor faced with a set of different models is not interested in selecting a best model but to convey all the available information to forecast the $t+1$ excess return and at the same time have a measure of the risk or uncertainty surrounding this forecast. Only at this point the investor can solve his own asset allocation problem. Since any model will only be an approximation to the generating mechanism and in many economic applications misspecification is inevitable, of substantial consequence and of an intractable nature, the strategy of choosing only the 'best' model (i.e. thin modelling) seems to be rather restrictive. If the economy features a wide-spread, slowly moving component that is approximated by an average of many variables through time but not by any single economic variable, then models that concentrate on parsimony could be missing it.

Furthermore if the true process is sufficiently complex, then the reduction strategy can lead to a model ('best' according to some criterion) which is more weakly correlated of the true model than the combination of different models.

In this paper we propose a novel methodology which extends the proposal contained in the original paper by P&T to deal explicitly with model uncertainty. The remainder of the paper is organized as follows: Section 2 discusses our proposal to deal with model uncertainty under rich parameterization for the predictive models. Section 3 re-assesses the original evidence on the statistical and economic significance of the predictability of stock returns by extending the data-set to the nineties and by evaluating comparatively alternative modelling strategies. Then we assess the statistical and economic significance of the predictions through a formal testing procedure and their use in a trading strategy. The last section concludes by providing an assessment of our main findings.

3.2 Recursive modelling: thin or thick ?

3.2.1 Thick Modelling

P&T (1995) consider the problem of an investor allocating his portfolio between a safe asset denominated in dollar and US stocks. The decision on portfolio allocation is then completely determined by the forecast of excess returns on US stock. Their allocation strategy is such that the portfolio is always totally allocated into one asset, which is the safe asset if predicted excess returns are negative, and shares if the predicted excess returns are positive. The authors forecast excess US stock returns by concentrating on an established benchmark set of regressors over which they conduct the search for a "satisfactory" predictive model. They focus on modelling the decision in real-time. To this end they implement a recursive modelling approach, according to which at each point in time, t, a search over a base set of observable k regressors is conducted to make one-period-ahead forecast. In each period they estimate a set of regressions spanned by all the possible permutations of the k regressors. This gives a total of 2^k different models for excess return. Models are estimated recursively, so that the dataset is expanded by one observation in each period. Therefore, a total of 2^k models are estimated in each period from 1959:12 to 1992:11 to generate a portfolio allocation.

P&T estimate all the possible specifications of the following forecasting equation:

$$(x_{t+1} - r_{t+1}) = \beta_i' \mathbf{X}_{t,i} + \varepsilon_{t+1,i} \tag{3.0}$$

where x_{t+1} are the monthly returns on the S&P500 Index and r_{t+1} are the monthly returns on the US dollar denominated safe asset (1-month T-bill), $\mathbf{X}_{t,i}$ is the set of regressors, observable at time t, included in the i-th specification ($i = 1, \ldots 2^k$) for the excess return. The relevant regressors are chosen from a benchmark set containing the dividend yield YSP_t, the price-earnings ratio PE_t, the 1-month T-bill rate $I1_t$ and its lag $I1_{t-1}$, the 12-month T-bill rate $I12_t$ and its lag $I12_{t-1}$, the year-on-year lagged rate of inflation π_{t-1}, the year-on-year lagged change in industrial output ΔIP_{t-1}, and the year-on-year lagged growth rate in the narrow money stock ΔM_{t-1}. A constant is always included and all variables based on macroeconomic indicators are measured by 12-month moving averages to decrease the impact of historical data revisions on the results[3].

At each sample point the investor computes OLS estimates of the unknown parameters for all possible models, chooses one forecast for excess returns given the predictions of $2^k = 512$ models and maps this forecast into a portfolio allocation by choosing shares if the forecast is positive and the safe asset if the forecast is negative. P&T select in each period only one forecast, i.e. the one

[3] See our Data Appendix for further details.

generated by the best model selected on the basis of a specified selection criteria which weights goodness of fit against parsimony of the specification (such as adjusted R^2, BIC, Akaike, Schwarz). We follow Granger (2003) and label this approach 'thin' modelling in that the forecast for excess returns and consequently the performance of the asset allocation are described over time by a thin line.

The specification procedure mimics a situation in which variables for predicting returns are chosen in each period from a pool of potentially relevant regressors according to the behavior often observed in financial markets of attributing different emphasis to the same variables in different periods. Obviously, keeping track of the selected variables helps the reflection on the economic significance of the 'best' regression.

The main limitation of thin modelling is that model, or specification, uncertainty is not considered. In each period the information coming from the discarded $2^k - 1$ models is ignored for the forecasting and portfolio allocation exercise.

This choice seems to be particularly strong in the light of the results obtained by Bayesian research, which stresses the importance of estimation risk for portfolio allocation[4]. A natural way to interpret model uncertainty is to refrain from the assumption of the existence of a "true" model and attach instead probabilities to different possible models. This approach has been labelled 'Bayesian Model Averaging'[5]. Bayesian methodology reveals the existence of in sample and out of sample predictability of stock returns, even when commonly adopted model selection criteria fail to demonstrate out of sample predictability.

The main difficulty with the application of Bayesian Model Averaging to problems like ours lies with the specification of prior distributions for parameters in all 2^k models of our interest.

[4]See, for example, Barberis (2000), Kandel and Stanbaugh (1996).

[5]For recent surveys of the literature about Bayesian Model Selection and Bayesian Model Averaging see respectively Chipman et al. (2001) and Hoeting et al. (1999). Avramov (2002) provides an interesting application.

Recently, Doppelhofer et al. (2000) have proposed an approach labelled 'Bayesian Averaging of Classical Estimates'(BACE) which overcomes the need of specifying priors by combining the averaging of estimates across models, a Bayesian concept, with classical OLS estimation, interpretable in the Bayesian camp as coming from the assumption of diffuse, non-informative, priors.

In practice BACE averages parameters across all models by weighting them proportionally to the logarithm of the likelihood function corrected for the degrees of freedom, using then a criterion similar to the Schwarz model selection criterion. It is important to note that the consideration of model uncertainty in our context generates potential for averaging at two different levels: averaging across the different predicted excess returns and averaging across the different portfolio choices driven by the excess returns.

There is also a vast literature[6] about forecast combination showing that combining in general works.

All forecasting models can be interpreted as a parsimonious representations of a General Unrestricted Model (GUM). Such approximations are obtained through the reduction process, which shrinks the GUM towards the local DGP (LDGP)[7]. White has shown that if the LDGP$\subset$GUM, then asymptotically the reduction process converges to the LDGP. However, there is the possibility that the LDGP is only partially contained in the GUM or completely outside the GUM. In this case the reduction procedure will converge asymptotically to a model that is closest to the true model, according to some distance function. As pointed out by Granger and Jeon (2003) there are good reasons for thinking that the thin modelling approach may not be a good strategy because a remarkable amount of information is lost. There are also a few recent results

[6] An incomplete list includes Chan-Stock-Watson (1999), Clemen (1989), Diebold-Pauly (1987), Elliott-Timmermann (2002), Giacomini and White (2002), Granger (2002), Clements and Hendry (2001), Marcellino (2002), Stock and Watson (2001,2003).

[7] An overview of the literature, and the developments leading to general-to-specific (Gets) modelling in particular, is provided by Campos, Hendry and Krolzig (2003).

(Stock and Watson (1999), Giacomini and White (2003)) suggesting that some important features of the data, as measured in term of forecast ability, can be lost in the reduction process. In fact, if the true DGP is quite complex, then the reduction process can lead to a model ('best' model) which contains less of the true model than the combination of different models. As pointed out by Granger (2003) it seems the economy might contain a wide-spread, slowly moving component that is approximated by an average of many variables through time but not by any single, economic variable, like a slow swing in the economy. If so, models that concentrate on parsimony could be missing this component.

This simple insight motivates the pragmatic idea of forecast combination, in which forecasts based on different models are the basic object of analysis. Forecast combination can be viewed as a key link between the short-run, real-time forecast production process, and the longer-run, ongoing process of model development. Furthermore in a large study of structural instability, Stock and Watson (1996) report that a majority of macroeconomic time series models undergo structural change, suggesting another argument for not relying on a single forecasting model. Finally another advantage of this approach is that a process, potentially non-linear, is linearized by looking at the linear specifications as Taylor expansions around different points.

The explicit consideration of estimation risks naturally generates 'thick' modelling, where both the prediction of models and the performance of the portfolio allocations over time are described by a thick line to take account of the multiplicity of models estimated. The thickness of the line is a direct reflection of the estimation risk.

Pesaran and Timmermann show that thin modelling allows to out-perform the buy and hold strategy. Re-evaluating their results from a thick modelling perspective raises immediately one question: *"why choose just one model to forecast excess returns?"*. In the next section we re-assess the evidence in P&T by using three different testing procedures of the performance of various forecasting models. We provide an empirical evaluation of the comparative performance of thin and thick modeling and address the issue of how to convey all the available information into a trading rule.

3.3 A first look at the empirical evidence

We start be replicating[8] the exercise in P&T by using the same dataset and by extending their original sample to 2001, keeping track of all the forecasts produced by taking into account the 2^k-1 combinations of regressors in a predictive model for US excess returns (the time-series of this variable is reported in Figure 1). We do so by looking at the within-sample econometric performance, at the out-of-sample forecasting performance and at the performance of the portfolio allocation.

Figure 2 allows us to analyze the within sample econometric performance by reporting the $\overline{R}^2$ for 2^k models estimated recursively. The difference in the selection criterion across different models is small, and almost negligible for models ranked close to each other.

We assess the forecasting performance of different models by using three type of tests: the Pesaran-Timmermann (1995) sign test, the Diebold-Mariano (1995) test and the White (2000) reality check. All tests and their implementation are fully described in an Appendix. The P&T sign test is an out-of-sample test of predictive ability, based on the proportion of times that the sign of a given variable is correctly predicted by the sign of some predictor. The Diebold-Mariano (1995) test is testing the null of a zero population mean loss differential between two forecasts. We use this test to evaluate the forecasting performance of thin modelling against several thick modelling alternatives. Finally, we implement the bootstrap reality check by White (2000), based

[8] In fact, we replicate the allocation results in the case of no transaction costs. Transaction costs do not affect the portfolio choice in the original exercise, therefore they do not affect the mapping from the forecasts to the portfolio allocation, which is the main concern of our paper.

on the consistent critical values given by Hansen (2001), to test the null that our benchmark (thin) model performs better than other available forecasting (thick) models. Importantly this testing procedure, allows us to take care of the possibility of data-snooping. We report the outcomes of the tests applied to the recursive modelling proposed by P&T in Table 1. We consider the whole sample 1959-2001 and we also split it into four decades. We compare the thin modelling, labelled as best (in terms of its adjusted R^2) with several thick modelling alternatives. We label *top x per cent*, the forecast obtained by averaging over the top x per cent models, ranked accordingly to their adjusted R^2. The line labelled *All* contains the results of averaging across all 2^k models. We then label *Median* the forecast obtained by considering the median of the empirical distribution of the within-sample performance. Lastly, we consider in the line *Dist* a synthetic measure of the skewness of this empirical distribution; in this case the selected prediction is that indicated by the majority of the models considered, independently from their ranking in terms of the within-sample performance. In general all tests show that it is possible to improve on the performance of the best model in terms of R^2 by using the information contained in the $2^k - 1$ models dominated (in many cases marginally) in terms of R^2. The sign test for the full sample shows that the thin modelling is always dominated by some thick modelling alternative. When different decades are considered, we observe that the percentage of correctly predicted signs is always significant for thick modelling in the three decades 60-70, 70-80 and 80-90, while the thin modelling alternative does not deliver a statistically significant value in the decade from 1980 to 1990. Interestingly, the decade 1990-2000 is an exception in that none of the strategies adopted delivers a statistically significant predictive performance. The evidence of the P&T tests is confirmed by the Diebold and Mariano tests. All the observed value for the statistics implemented on the full sample are negative and significant, showing that the null of equal predictive ability of thin and thick modelling is rejected, at one per cent level, independently form the adopted thick modelling specification. Such evidence is considerably weakened when the sample is split into decades. Finally the reported p-values for the

White reality check, show that the null that all the alternative thick modelling strategies are not better than the thin model is consistently rejected when the full sample is considered. Splitting the sample into decades weakens the results only for the period 1990-2000.

The results of the forecasting performance are confirmed by the performance of the portfolio allocation. We report in Figure 3 the cumulative end-of period wealth delivered by the portfolios associated with all 512 possible models, ranked in terms of their $\overline{R}^2$. Following P&T, portfolios are always totally allocated into one asset, which is the safe asset if predicted excess returns are negative, and shares if the predicted excess returns are positive. We add as a benchmark the final wealth given by the buy-and-hold strategy. Figure 3 shows that in general the value of the end-of-period wealth is not a decreasing function of the $\overline{R}^2$, and that the buy and hold strategy is in general dominated, again with the notable exception of the decade 1990-2000, where the buy and hold strategy gives the highest wealth.

To sum up, our evidence suggests that thick modeling dominates thin modelling but also that the evidence for excess return predictability is considerably weaker in the period 1990-2000[9]. In fact, over this sample, the adjusted R^2 of all models decreases substantially, the sign tests for predictive performance are not significant anymore, and the econometric performance-based portfolio allocation generate lower wealth than the buy-and-hold strategy.

In the next section we shall evaluate refinements in the specification and the modelling selection strategy in the spirit of thick modelling.

[9]This is also observed by Paye-Timmermann(2002).

3.4 Our proposal for thick modelling

In the light of the evidence reported in the previous section we propose extensions of the original methodology both at the stage of model specification and of portfolio allocation.

The empirical evidence reported in the previous section shows clearly that the ranking of models in terms of their within-sample performance does not match at all the ranking of models in terms of their ex-post forecasting power. This empirical evidence points clearly against BACE using within sample criteria to weight models. Consistent with this evidence, we opted for the selection method proposed by Granger and Yeon (2003) of using a '*... procedure [which] emphasizes the purpose of the task at hand rather than just using a simple statistical pooling...*' Our task at hand is asset allocation.

3.4.1 Model specification

At the stage of model specification we consider two issues: the importance of balanced regressions and the optimal choice of the window of observations for estimation purposes.

A regression is balanced when the order of integration of the regressors matches that of the dependent variables. Excess returns are stationary, but not all variables candidate to explain that are stationary. To achieve a balanced regression in this case, cointegration among the included non-stationary variables is needed. As shown by Sims, Stock and Watson (1990) the appropriate stationary linear combinations of non-stationary variables will be naturally selected by the dynamic regression, when all non stationary variables potentially included in a cointegrating relation are included in the model. Therefore, when model selection criteria are applied, one must make sure that such criteria do not lead to exclude any component of the cointegrating vector from the regression. Following Pesaran and Timmermann (2001) we divide variables in focal, labelled A_t

and secondary focal, labelled B_t. Focal variables are always included in all models, while the variables in B_t are subject to the selection process. We take these variables as those defining the long-run equilibria for the stock market. Following the lead of traditional analysis[10] and recent studies (Lander et al. (1997)) we have chosen to construct an equilibrium for the stock market by concentrating on a linear relation between the long term interest rates, R_t, and the logarithm of the earning price ratio, *ep*. Also recent empirical analysis (see Zhou, 1996) finds that stock market movements are closely related to shifts in the slope of the term structure. Such results might be explained by a correlation between the risk premia on long-term bonds and the risk premium on stocks. Therefore, we consider the term spread as a potentially important cointegrating relation. On the basis of this consideration we include in the set of focal variables the yield to maturity on 10-year government bonds (a variable which was not included in the original set of regressors in P&T), the log of the earning price ratio and the interest rate on 12-month Treasury Bills, to ensure that the selected model is balanced and includes the two relevant cointegrating vectors. We do not impose any restrictions on the coefficients of the focal variables.[11]

The second important issue at the stage of model selection is the choice of the window of observations for estimation (i.e. for how long a predictive relationship stays in effect)[12]. The question of stability is equally important since the expected economic value from having discovered

---

[10]

> *"...Theoretical analysis suggests that both the dividend yield and the earnings yield on common stocks should be strongly affected by changes in the long-term interest rates. It is assumed that many investors are constantly making a choice between stock and bond purchases; as the yield on bonds advances, they would be expected to demand a correspondingly higher return on stocks, and conversely as bond yields decline..."* (Graham and Dodd Security Analysis, 4th edition, 1962, p.510).

The above statement suggests that either the dividend yield or the earnings yield on common stocks could be used.

[11] We have assessed the choice of our focal variable by estimating recursively a VAR including the yield to maturity of 10-year government bonds, the log of the earning-price ratio and the interest rate on 12-month Treasury Bills. The null of no cointegration is always rejected when the Johansen (1995) procedure is implemented by allowing for an intercept in the cointegrating vectors. We choose not to impose any restriction on the number of cointegrating vectors and on cointegrating parameters as they are not constant over time (a full set of empirical results is available upon request).

[12] Recent empirical studies cast doubt upon the assumed stability in return forecasting models. An incomplete list includes Ang and Bekaert (2001), Lettau and Ludvigson (2001), Paye and Timmermann (2002).

a good historical forecasting model is much smaller if there is a high likelihood of the model breaking down subsequently.

In the absence of breaks in the DGP the usual method for estimation and forecasting is to use an expanding window. In this case, by augmenting an already selected sample period with new observations, more efficient estimates of the same fixed coefficients are obtained by using more information as it becomes available. However, if the parameters of the regression model are not believed to be constant over time, a rolling window of observations with a fixed size is frequently used. When a rolling window is used, the natural issue is the choice of its size. This problem has been already observed by Pesaran and Timmermann (2002) who provide an extensive analysis of model instability, structural breaks, and the choice of window observations. In line with their analysis we deal with the problem of window selection by starting from an expanding window, every time a new observation is available we run a backward CUSUM and CUSUM squared test to detect instability in the intercept and/or in the variance. We then keep expanding the window only when the null of no structural break is not rejected. Consider a sample of T observations and the following model:

$$y_{t,T} = \beta^{i'} x^i_{t,T} + u_{t,T} \; i = 1, ..., 2^k$$

where $y_{t,T} = (y_t, y_t, y_{t+2}, ..., y_T)$ and $x^i_{t,T} = \left(x^i_t, x^i_{t+1}, x^i_{t+2}, ..., x^i_T\right)$ where $T - t + 1$ is the optimal window and T the last available observation. Recall that we are interested in forecasting y_{T+1} given $x_{T+1,}\widehat{\beta}^{i'}$. The problem of the optimal choice of t given model i, can be solved by running a CUSUM test with the order of the observations reversed in time starting from the m-th observation and going back to the first observation available (we refer to this procedure as ROC). Critical values by Brown et al (1975) can be used to decide if a break has occurred. Unlike the Bai-Perron method, the ROC method does not consistently estimate the breakpoint[13]. On the other hand, the simpler

[13]As pointed out by Pesaran and Timmermann (2002), ironically this may well benefit the ROC method in the context of forecasting since it can be optimal to include pre-break data in the estimation of a forecasting model. Although doing so leads to biased predictions, it also reduces the parameter estimation uncertainty.

look-back approach only requires detecting a single break and may succeed in determining the most recent breakpoint in a manner better suited for forecasting. Once a structural break (either in the mean or in the variance) has been detected, we have found the appropriate t. Clearly the appropriate t can be the first observation in the sample (in this case we have an expanding window) or any number between 1 and m (flexible rolling window). This procedure allows us to optimally select the observation window[14] for each of the 2^k different models estimated at time t.

In terms of model selection we have now several methodologies available: the original P&T recursive estimation (based on an expanding window of observations) with no division of variables in focal and semi-focal, the rolling estimation (based on a fixed window of sixty observations) with no division of variables into focal and semi-focal, the balanced recursive estimation, in which variables are divided into focal and non-focal, to make sure that cointegrating relationship(s) are always included in the specification, and a flexible estimation, in which the optimal size for the estimation window is chosen for all possible samples. We consider two versions of the flexible estimation that differ by the division of variables into focal and semi-focal.

3.4.2 Asset Allocation

To analyze how the value of the investor's portfolio evolves through time, we first introduce some notations. Let W_t be the funds available to the investor at the end of period t, α_t^S the numbers of shares held at the end of period t, r_t^s the rate of return on S&P500 and r_t^b, the rate of return on safe asset in period t, S_t and B_t the investor's position in stock and safe assets at the end of period t, respectively. At a particular point in time, t, the budget constraint of the investor is given by:

$$W_t = (1 + r_{t-1}^s)S_{t-1} + (1 + r_{t-1}^b)B_{t-1}$$

[14] We impose that the shortest observation window automatically selected cannot be smaller than 2 or 3 times the dimension of the parameters' vector. So also the minimum observation window is a function of regressors included in each of 2^k different models.

P&T propose an allocation strategy such that the portfolio is always totally allocated into one asset, which is the safe asset if predicted excess returns are negative, and shares if the predicted excess returns are positive. We consider three alternative ways of implementing thick modelling when allocating portfolios. Given the 2^k forecasts for excess returns in each period define α_t^S and $\alpha_t^B = \left(1 - \alpha_t^S\right)$ to be respectively the weight on stocks and the safe asset (short term bills), let $\{y_i\}_{i=1}^{2^k}$ the full set of excess return forecasts obtained in the previous step, and let $n = \omega' 2^k$, where $\omega = [.01, .05, .1, .2, .3, .4, .5, .6, .7, .8, .9, 1]$ is the set of weights, in terms of the percentage of the model ordered according to their adjusted R^2, chosen to build up the appropriate trimmed means of the available forecasts. Then we propose the following allocation criteria:

Distribution-Thick-Modelling: We look at the empirical distribution of the forecasts to apply the following criterion:

1. $\alpha_{\omega_i}^S = \begin{cases} 1 & if \quad \left[\frac{\sum_{i=1}^{n_{\omega_j}} (y_i > 0)}{n_{\omega_i}}\right] > 0.5 \\ 0 & otherwise \end{cases}$;

   where $n_{\omega_j}(y_i > 0)$ is the number of models giving a positive prediction for excess returns within the j-th class of the trimming grid (for example $n_{\omega_2}(y_i > 0)$ is the number of models in the best 5 per cent of the ranking in term of their $\overline{R}^2$ predicting a positive excess return). In practice if more than 50 percent of the considered models predict an upturn (downturn) of the market, we put all the wealth in the stock market (safe asset).

2. *Meta-Thick-Modelling:* We use the same criterion as above, to derive a less aggressive portfolio allocation, in which corner solution are the exceptions rather than the rule:

   $\alpha_{\omega_i}^S = \left[\frac{\sum_{i=1}^{n_{\omega_i}} (y_i > 0)}{n_{\omega_i}}\right]$;

3. *Kernel-Thick-Modelling*: we compute the weigthed average of predictions $\overline{y}$ (with weights based on the relative adjusted-R^2, through a triangular kernel function that penalizes deviations from the best model in terms of R^2 and the bandwidth determined by the number

of observations) and then we apply this rule:

$$\alpha^S_{\omega_i} = \begin{cases} 1 & if \ \ \overline{y} > 0 \\ 0 & otherwise \end{cases}.$$

3.5 Empirical Results

Our empirical results are reported in Table 2-4 and Figures 3-5.

In Tables 2-4 we evaluate the forecasting performance of all methodologies by using our three testing procedure.

In Table 2 we report the results of the Pesaran-Timmermann market-timing test of thin and thick modelling excess return forecasts, in Table 3 we report the results of the Diebold-Mariano test of equal predictive ability between thin and thick modelling excess returns forecasts, finally in Table 4 we report the results for White's reality check to test the null that thin modelling based forecasts out-perform thick modelling based forecasts.

Overall, all three tests suggest that the flexible estimation delivers the best results. The most remarkable improvements occurs when the Diebold-Mariano and White's reality check are implemented over the decade 1990-2000. The P&T sign test confirms the results of the other two tests but also signals that the null that any chosen predictor has no power in predicting excess returns over the decade 1990-2000 cannot be rejected.

On the basis of this evidence we proceed to evaluate the performance of asset allocation based on thin and thick modelling, considering the buy-and-hold strategy as a benchmark.

Figures 4-5 evaluate the performance of different portfolio allocation criteria, by comparing the end-of-period cumulative wealth associated with the recursive estimation and the rolling estimation with optimally chosen window and focal regressors with the cumulative wealth associated with a simple buy-and-hold strategy[15]. Each figure considers an estimation criteria and reports the performance of portfolio allocations for the thin modeling approach and different types of thick modelling along with the buy-and-hold strategy. We report, for the full sample and for the four decades, the end of period wealth associated with a beginning of period wealth of 100.

With very few exceptions thick modelling dominates thin modelling. Moreover, the more articulated model specification procedures deliver better results than the simple recursive criterion. The best performance is achieved when the distribution-thick modelling is applied to the best 20 per cent of models in terms of their adjusted R^2. Model-based portfolio allocations dominate the buy-and-hold strategy over the whole sample and in the decades 1970-80 and 1980-90. More complicated specification procedures tend to give a weaker out-performance relative to the buy-and-hold than the simple recursive specification. The evidence for the decade 1960-70 is mixed in the sense that not all econometric based strategies dominate on buy-and-hold strategy. In the last decade the buy-and-hold strategy is never out-performed, however the dominance of thick modelling over thin modelling becomes stronger.

3.6 Conclusions

In this paper, we have reassessed the results on the statistical and economic significance of the predictability of stock returns provided by Pesaran and Timmermann (1995) for US data to propose a novel approach for portfolio allocation based on econometric modelling. We find that the results based on the thin modelling approach originally obtained for the sample 1960-1992 are considerably weakened in the decade 1990-2000.

We then show that the incorporation of model uncertainty substantially improves the

[15] Evaluation has been also conducted in terms of period returns and Sharpe-ratios, results are available upon request.

performance of econometric based portfolio allocation.

The portfolio allocation based on a strategy giving weights to a number of models rather than to just one model leads to systematic over-performance of portfolio allocations among two assets based on a single model. However, even thick modelling does not guarantee a constant over-performance with respect to a typical market benchmark for our asset allocation problem. To this end we have observed that combining thick modelling with a model specification strategy that imposes balanced regressions and chooses optimally the estimation window reduces the volatility of the asset allocation performance and delivers a more consistent out-performance with respect to the simple buy-and-hold strategy.

3.7 Appendix

3.7.1 Data Appendix

In the Pesaran-Timmermann (1995) dataset (PT95) the data sources were as follows: stock prices were measured by the Standard &Poor's 500 index at close on the last trading day of each month. These stock indices, as well as a monthly average of annualized dividends and earnings, were taken from Standard & Poor's Statistical Service. The 1-month T-bill rate was measured on the last trading day of the month and computed as the average of the bid and ask yields. The source was the Fama-Bliss risk free rates file on the CRSP tapes. The same for 12-month discount bond rate. The inflation rate was computed using the producer price index for finished goods from Citibase, and the rate of change in industrial production was based on a seasonally adjusted index for industrial production (Citibase). The monetary series were based on the narrow monetary aggregates published by the Fed of St. Louis and provided by Citibase.

The extended dataset has been obtained merging P&T original dataset (1954.1-1992.12) with new series retrived from DATASTREAM and FRED for the sample 1993.1-2001.9. All the financial variables are measured on the last trading day of each month.

	Code	Description
$P_t^{stock,US}$	TOTMKUS(RI)	US-DS MARKET - TOT RETURN IND
dy_t^{US}	TOTMKUS(DY)	US -DS market- Dividend yield
pe_t^{US}	TOTMKUS(PE)	US-DS MARKET - PER
$r1_t^{US}$	ECUSD1M	US EURO-$ 1 MONTH (LDN:FT) - MIDDLE RATE
ppi_t^{US}	USOCPRODF	US PPI - MANUFACTURED GOODS NADJ
$r12_t^{US}$	ECUSD1Y	US EURO-$ 1 YEAR (LDN:FT) - MIDDLE RATE
ip_t^{US}	USINPRODG	US INDUSTRIAL PRODUCTION
$M0_t^{US}$	USM0....B	US MONETARY BASE CURA
$R10Y_t^{US}$	BMUS10Y(RY)	US YIELD-TO-MATURITY ON 10_YEAR GOV.BONDS

The data are available in Excel format from the following website: http://www.igier.unibocconi.it/favero (section working papers)

3.7.2 Testing Performance of Various Forecasting Models

In this paper we focus on out-of-sample tests of stock predictability. Out-of-samples tests are more stringent than in-sample tests and have important advantages over in sample tests in assessing the predictability of stock returns. We analyze out-of-sample predictive ability using 3 recently developed statistics.

The first one is the market timing test proposed by Pesaran and Timmermann (1992). The sign test is based on the proportion of times that the sign of a given variable y_t is correctly predicted in the sample by the sign of the predictor x_t . Under the null hypothesis that x_t has no power in predicting y_t the proportion of times that the sign is correctly predicted has a binomial distribution with known parameters, therefore a test of the null of predictive failure is constructed by comparing the observed proportion of sign correctly predicted with the proportion of sign correctly predicted under the null. The test statistic is computed as

$$Sn = \frac{P - P^*}{\{V(P) - V(P^*)\}^{1/2}} \sim N(0,1)$$

where:

$$P = \bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$$

$$P^* = P_y P_x + (1 - P_y)(1 - P_x)$$

$$V(P^*) = \frac{1}{n} P^* (1 - P^*)$$

$$V(P) = n \left(\begin{array}{c} (2P_y - 1)^2 P_x (1 - P_x) + (2P_x - 1)^2 P_y (1 - P_y) + \\ + \frac{4}{n} P_y P_x (1 - P_y)(1 - P_x) \end{array} \right)$$

Z_i is an indicator variable which takes value of one when the sign of y_t is correctly predicted by x_t, and zero otherwise, P_y is the proportion of times y_t takes a positive value, P_x is the proportion of times x_t takes a positive value.

The second one is the popular Diebold and Mariano (1995) statistic for equal predictive accuracy where we are testing the null hypothesis of a zero population mean loss differential between two forecasts. This test has a standard limiting distribution when comparing forecasts from non-nested models. However we are comparing forecasts from nested models, so we follow the raccomendation of Clark and McCracken (2001b) and base our inference on a boostrap procedure similar to one used in Kilian (1999). In order to derive the correct distribution for the statistic we apply the bootstrap in the following way. Let $d_{k,t}$, $t = 1, ..., n$ be the sequence of the realized difference in loss between model k and a benchmark model.

1. run the regression $E(d_t) = c + \epsilon_t$;
2. compute $\widehat{\epsilon}_t$ and generate B bootstrap samples[16];
3. generate B bootstrap responses $E(d_t)^{*1}, ..., E(d_t)^{*B}$ according to $E(d_t)^{*b} = \widehat{c} + \widehat{\epsilon}_t^{*b}$;
4. the new bootstrap dataset is given by $\left(E(d_t)^{*b}, c\right)$;
5. compute the t-value of the constant and denote it by t^{*b};
6. derive the distribution of t^{*b};
7. compute p-value as $\#\left(t^{actual} > t^{*b}\right)/B$.

The third procedure we implement is the Bootstrap Reality Check by White (2000) with the consistent values given by Hansen (2001). In this case we are testing the null that a model

[16]There are different ways to generate the resamples: one approach is the stationary bootstrap by Politis and Romano (1994), another is the block bootstrap of Kunsch (1989).

(benchmark) performs better than other avaiable forecasting models in a given sample, taking care of data snooping. The need to test for Superior Predictive Ability arises from a situation in which, like our case, a family of forecasting models are compared in terms of their predictive ability defined in the form of a loss function. The question of interest is whether any alternative model is a better forecasting model is a better forecasting model than a benchmark model. When a large number of models are investigated prior to the selection of a model, then the search over models must be taken into account when making inference. After a search over several models, the relevant question is whether the excess performance of an alternative model is significant or not.

Let $X_k(t), t = 1, ..., n$ be the sequence of realized performance of model k relative to a benchmark, $k = 0, ..., M$.

Let $b = 1, ..., B$ index the resamples of $\{1, .., n\}$, given by $\theta_b(t), t = 1, ..., n$ where B denotes the number of bootstrap resamples generated by the stationary boostrap of Politis and Romano (1994). The b'th bootstrap resample is defined as: $X^*_{k,b}(t) = X_k(\theta_b(t)) - g(\overline{X}_{n,k}), b = 1, .., B, t = 1, ..., n$ where $g(x) = \begin{cases} 0 & if\ x \leq -A_{n,k} \\ x & otherwise. \end{cases}$ where $A_{n,k}$ is a correction factor depending on an estimate of $var\left(n^{1/2}\overline{X}_{n,k}\right)$. For $b = 1, ..., B$, we calculate $\overline{X}^*_{n,\max,b} = \max_k \overline{X}^*_{n,k,b}$, and th estimated p-value is given by

$$\widehat{p} = \sum_{b=1}^{B} \frac{1\left(\overline{X}^*_{n,\max,b} > \overline{X}_{n,\max}\right)}{B}.$$

In both cases is very important to specify the loss function we have in mind. Evaluation of forecasting skills of a forecast producer may be best carried out using one of the purely statistical measures, while for a user forecast evaluation requires a decision based approach[17]. From a user's perspective forecast accuracy is best judged by its expected economic value, the characterization of

[17]Whittle notes 'Prediction is not an end in itself, but only a means of optimizing current actions against the prospect of an uncertain future'. To evaluate forecasts we need to know how and by whom forecasts are used. See Pesaran and Skouras (2002) for further details.

which requires a full specification of the user's decision environment. In our case, where the objective of forecasting is relatively uncontroversial, the importance of economic measures of forecast accuracy has been widely acknowledged and is straightforward. However, since we report economic measures of forecast accuracy in the next section, where we discuss the asset allocation performance, we decide to use the standard MSE loss function to test the different forecasts.

Bibliography

Aiolfi M, Favero CA, Primiceri G. 2001. Recursive 'Thick' Modelling of Excess Return and Dynamic Portfolio Allocation. IGIER Working Paper N. 197.

Ait-Sahalia Y, Brandt M. 2001. Variable Selection for Portfolio Choice. *Journal of Finance* **56**: 1297-1351.

Ang A, Bekaert G. 2002. Stock Return Predictability: Is it There?. Working Paper, Columbia Business School.

Avramov D. 2002. Stock return predictability and model uncertainty. *Journal of Financial Economics,* **64**:423-458.

Barberis N. 2000. Investing for the Long run when Returns are Predictable. *Journal of Finance* **55**, 1: 225-264

Bossaerts P, Hillion P. 1999. Implementing statistical criteria to select return forecasting models: what do we learn?. *The Review of Financial Studies* **12**: 405-428.

Brandt M. 1999. Estimating portfolio and consumption choice: A conditional Euler equations approach. *Journal of Finance* **54**: 1609–1646.

Campbell JY, Shiller R. 1987. Cointegration and tests of present value models. *Journal of Political Economy* **95**: 1062-1088.

Campbell JY, Shiller R. 1988a .The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies* **1**: 195-227.

Campbell JY, Shiller R. 1988b. Stock prices, earnings, and expected dividends. *Journal of Finance* **43**: 661-676.

Campbell JY, Lo A, McKinlay. 1997. *The Econometrics of Financial Markets*. Princeton

University Press.

Campbell JY, Viceira L. 1999. Consumption and portfolio decisions when expected returns are time-varying. *Quarterly Journal of Economics* **114**: 433-495.

Campos J., Hendry D.F., Krolzig H.-M.. 2003 Consistent Model Selection by an Automatic Gets Approach, *Oxford Bulletin of Economics & Statistics.*, **65**: 803-820

Chan, Stock J, Watson M. 1999. A Dynamic Factor Model Framework for Forecast Combination. *Spanish Economic Review* **1**: 91-121.

Chipman H, George EI, McCulloch RE. 2001. The practical implementation of Bayesian model selection. *IMS Lecture Notes - Monograph Series* Vol **38**.

Clemen RT. 1989. Combining Forecasts: A Review and Annotated Blibliography. *International Journal of Forecasting* **5**: 559-581.

Clements MP, Hendry DF. 2001. *Forecasting Economic time-Series.* Cambridge University Press, Cambridge.

Cochrane J. 1999. Portfolio advice for a multifactor world. NBER wp 7170.

Diebold F, Mariano R. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**: 253-265.

Diebold F., Pauly P. 1987. Structural Change and the Combination of Forecasts, *Journal of Forecasting,* **6**:21-40.

Doppelhofer G, Miller RI, Sala-i-Martin. 2000. Determinants of long-term growth: a Bayesian averaging of classical estimates (BACE) approach. NBER wp 7750.

Elliott G, Timmerman A. 2002. Optimal Forecast Combinations under General Loss Functions and Forecast Error Distributions. UC San Diego Department of Economics Discussion

Paper 03-09.

Fama E, French KR. 1988. Dividend yields and expected stock returns. *Journal of Financial Economics* **22**: 3–25.

Giacomini R, White H. 2003. Test of Conditional Predictive Ability. UC San Diego Department of Economics Discussion Paper 03-09.

Granger CWJ, Newbold P. 1986. *Forecasting Economic Time Series*, 2nd edn. San Diego Academic Press.

Granger CWJ, Timmermann A. 1999. Data mining with local model specification uncertainty: a discussion of Hoover and Perez. *The Econometrics Journal* **2**: 220-226.

Granger CWJ, Jeon Y. 2003. Thick modeling. Mimeo, UC San Diego Department of Economics.

Granger CWJ. 2003. Modeling Conditional Distribution. Mimeo, UC San Diego Department of Economics.

Hansen PR. 2001. An Unbiased and Powerful Test for Superior Predictive Ability. Brown University Department of Economics Discussion Paper 01-06.

Hoeting J, Madigan D, Raftery A, Volinsky C. 1999. Bayesian Model Averaging: a Tutorial. Technical Report 9814, Department of Statistics, Colorado State University.

Johansen S. 1995. Likelihood-based inference cointegrated vector auto-regressive models. Oxford University Press.

Kandel S, Stambaugh RS. 1996. On the Predictability of Stock Returns: An Asset-Allocation Perspective. *Journal of Finance* **51**, 2: 385-424.

Keim D, Stambaugh RS. 1986. Predicting returns in the stock and bond markets. *Journal* *of Financial Economics* **17**: 357-390

Lander J, Orphanides A, Douvogiannis M. 1997. Earning forecasts and the predictability of stock returns: evidence from trading the S&P. Board of Governors of the Federal Reserve System. http:\\ www.bog.frb.fed.org.

Lamont O. 1998. Earnings and expected returns. *Journal of Finance* **53**, 5: 1563-1587

Lettau, Ludvigson S. 2001. Consumption, Aggregate Wealth and Expected Stock Returns. *Journal of Finance* **56**, 3: 815-849

Marcellino M. 2002. Forecast Pooling for Short Time Series of Macroeconomic Variables. IGIER Working Paper 212.

Paye B, Timmermann A. 2002. How stable are Financial Prediction Models? Evidence from US and International Stock Market. UC San Diego Department of Economics Discussion Paper 02-13.

Pesaran MH, Timmermann A. 1992. A simple non-parametric test of predictive performance. *Journal of Business and Economics Statistics* **10**: 461-465.

Pesaran MH, Timmermann A. 1995. Predictability of Stock Returns: Robustness and Economic Significance. *Journal of Finance* **50**, 4: 1201-1228.

Pesaran MH, Timmermann A. 2000. A Recursive Modelling Approach to Predicting UK Stock Returns. *The Economic Journal* **110**: 159-191.

Pesaran MH, Timmermann A. 2002. Market Timing and Return Prediction under Model Instability. *Journal of Empirical Finance* **9**: 495–510.

Rapach DE, Wohar ME. 2002. Financial Variables and the Predictability of Stock and Bond Returns: An Out-of-Sample Analysis. Mimeo.

Raftery A, Madigan D, Hoeting J. 1997. Bayesian model averaging for linear regression models. *Journal of the American Statistical Association* **92** (437): 179-191

Samuelson PA. 1969. Lifetime portfolio selection by dynamic stochastic programming. *Review of Economics and Statistics* **51**: 239-246.

Siegel J. 1994. *Stocks for the Long Run*. Richard D.Irwin, Burr Ridge, III.

Sullivan R, Timmermann A, White H. 1999. Data-snooping, technical trading rules performance and the bootstrap. *Journal of Finance* **54**: 1647-1692

Sims C, Stock J, Watson M. 1990. Inference in linear time-series models with some unit roots. *Econometrica* **58**: 113-144

Stock J, Watson M. 1999. A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series. In Engle, RF, White H., eds., *Cointegration, Causality, and Forecasting: a Festschrift in Honour of Clive W.J. Granger*, Chapter 1. Oxford University Press.

Zhou C. 1996. Stock market fluctuations and the term structure. Board of Governors of the Federal Reserve System. http: \\www.bog.frb.fed.org.

White H. 2000. A Reality Check for Data Snooping. *Econometrica* **68**: 1097-1126.

Table 3.1: Forecasting Performance of Thin Vs. Thick Modelling

The results are based on recursive least squares estimation with the constant term as the only focal variable. The Pesaran-Timmermann market timing test (PT) is the percentage of times that the sign of the realized excess returns is correctly predicted by the forecast combination strategy reported by rows. The Diebold and Mariano (DM) test statistic is used to test the null of equal predictive ability between thin and different versions of thick modelling. The White Bootstrap Reality Check (RC) is used to test the null that the in-sample best model performs better than all the other available forecasting models. **,* indicate significance at the 1%, 5% levels, respectively. For White Bootstrap Reality Check we report the p-value.

	PT	DM	RC	PT	DM	RC
	Panel A: 1960-1970			Panel B 1970-1980		
Best	0.57			0.62**		
Top 1%	0.57	-1.20	0.00	0.62**	-0.73	0.00
Top 5%	0.56	-0.82	0.00	0.63**	-0.20	0.00
Top 10%	0.56	-1.08	0.00	0.63**	-0.24	0.00
Top 20%	0.56	-0.85	0.00	0.61**	-0.65	0.00
Top 30%	0.57	-1.03	0.01	0.63**	-0.58	0.01
Top 40%	0.58*	-1.04	0.03	0.60*	-0.83	0.03
Top 50%	0.59*	-1.13	0.03	0.60*	-0.99	0.04
Top 60%	0.58*	-1.19	0.06	0.60*	-0.98	0.06
Top 70%	0.58*	-1.14	0.07	0.61**	-1.08	0.07
Top 80%	0.58*	-1.02	0.10	0.60*	-1.07	0.10
Top 90%	0.58*	-0.96	0.13	0.59*	-1.00	0.12
All	0.57	-0.98	0.16	0.58*	-0.88	0.13
Median	0.57		0.14	0.60*		0.13
Dist	0.57		0.00	0.60*		0.00
	Panel C 1980-1990			Panel D 1990-2000		
Best	0.57			0.48		
Top 1%	0.57	1.11	0.00	0.49	0.33	0.12
Top 5%	0.58	-0.77	0.00	0.46	0.84	0.31
Top 10%	0.59	-1.31	0.00	0.46	1.51	0.39
Top 20%	0.60*	-1.28	0.00	0.47	1.81	0.42
Top 30%	0.62*	-1.43	0.02	0.46	1.85	0.42
Top 40%	0.64**	-1.34	0.03	0.47	1.68	0.41
Top 50%	0.64**	-1.33	0.05	0.49	1.44	0.41
Top 60%	0.64**	-1.32	0.06	0.48	1.11	0.40
Top 70%	0.64**	-1.31	0.07	0.48	0.89	0.39
Top 80%	0.63**	-1.29	0.08	0.48	0.62	0.39
Top 90%	0.62**	-1.22	0.09	0.47	0.26	0.41
All	0.62*	-1.16	0.11	0.47	-0.22	0.41
Median	0.62*		0.10	0.45		0.41
Dist	0.62*		0.00	0.45		0.00
	Panel E 1960-2001					
Best	0.56*					
Top 1%	0.56*	-1.67	0.00			
Top 5%	0.55*	-5.21**	0.00			
Top 10%	0.55*	-5.35**	0.00			
Top 20%	0.55*	-6.21**	0.00			
Top 30%	0.56**	-6.37**	0.00			
Top 40%	0.57**	-6.57**	0.00			
Top 50%	0.57**	-6.46**	0.01			
Top 60%	0.57**	-6.24**	0.01			
Top 70%	0.57**	-6.02**	0.01			
Top 80%	0.57**	-5.79**	0.01			
Top 90%	0.56**	-5.57**	0.02			
All	0.56**	-5.09**	0.03			
Median	0.55**		0.02			
Dist	0.55**		0.00			

Table 3.2: Pesaran-Timmermann Market Timing Test of Thin and Thick Modelling Excess Return Forecasts

Each Panel reports the proportion of times that in a given sample the sign of realized excess returns is correctly predicted by the sign of alternative thin and thick modelling one step ahead forecasts generated by five different estimation strategies. '**','*' indicate significant evidence of market timing at the 5%, and 10% levels, respectively. Top x% is the combination of the trimmed mean of the best x% forecasting models, Med is the combination scheme based on the median, and Dist is the combination scheme based on the majority rule applied to all the available forecasting models. REC, ROLL, OW denote recursive estimation, rolling estimation with fixed window length, optimal estimation window, respectively. The numbers in square brackets show the number of focal variables considered. [1] is just the constant, while [4] denotes the following set of regressors: constant, log of the price-earning ratio, yield-to maturity on long term bonds, yield on 12-month Treasury Bills

	REC [1]	ROLL [1]	REC [4]	OW [1]	OW [4]	REC [1]	ROLL [1]	REC [4]	OW [1]	OW [4]
			Panel A: 1960-1970					Panel B 1970-1980		
Best	0.57	0.55	0.53	0.50	0.54	0.62**	0.51	0.57	0.57	0.56
Top 1%	0.57	0.54	0.52	0.53	0.52	0.62**	0.52	0.58*	0.56	0.55
Top 5%	0.56	0.55	0.53	0.53	0.53	0.63**	0.52	0.57	0.57	0.57
Top 10%	0.56	0.57	0.53	0.51	0.53	0.63**	0.53	0.57	0.57	0.59*
Top 20%	0.56	0.57	0.53	0.54	0.54	0.61**	0.49	0.57	0.59*	0.54
Top 30%	0.57	0.55	0.55	0.53	0.54	0.63**	0.51	0.57	0.59*	0.55
Top 40%	0.58*	0.56	0.56	0.57	0.54	0.60*	0.53	0.54	0.62**	0.54
Top 50%	0.59*	0.56	0.55	0.57	0.54	0.60*	0.53	0.55	0.60*	0.54
Top 60%	0.58*	0.56	0.54	0.56	0.53	0.60*	0.54	0.56	0.61**	0.53
Top 70%	0.58*	0.57	0.57	0.57	0.52	0.61**	0.57*	0.57	0.61**	0.54
Top 80%	0.58*	0.57	0.53	0.56	0.53	0.60*	0.55	0.54	0.60*	0.54
Top 90%	0.58*	0.57	0.53	0.55	0.54	0.59*	0.56	0.54	0.60*	0.55
All	0.57	0.56	0.54	0.55	0.55	0.58*	0.56	0.55	0.59*	0.55
Median	0.57	0.53	0.55	0.57	0.55	0.60*	0.57	0.54	0.61**	0.53
Dist	0.57	0.53	0.55	0.57	0.55	0.60*	0.57	0.54	0.61**	0.53
			Panel C 1980-1990					Panel D 1990-2000		
Best	0.57	0.57*	0.59	0.57*	0.53	0.48	0.49	0.50	0.48	0.46
Top 1%	0.57	0.58*	0.60*	0.56*	0.54	0.49	0.51	0.51	0.47	0.47
Top 5%	0.58	0.59*	0.59	0.59*	0.57	0.46	0.52	0.50	0.45	0.49
Top 10%	0.59	0.60*	0.61*	0.60*	0.59*	0.46	0.53	0.50	0.44	0.47
Top 20%	0.60*	0.59*	0.59	0.61**	0.57	0.47	0.51	0.47	0.50	0.46
Top 30%	0.62*	0.60**	0.61*	0.60**	0.57	0.46	0.53	0.48	0.51	0.48
Top 40%	0.64**	0.61**	0.62*	0.60**	0.59*	0.47	0.54	0.48	0.47	0.46
Top 50%	0.64**	0.63**	0.62*	0.60**	0.59*	0.49	0.53	0.48	0.49	0.49
Top 60%	0.64**	0.61**	0.60*	0.57*	0.58	0.48	0.55	0.47	0.51	0.52
Top 70%	0.64**	0.63**	0.60*	0.60**	0.59*	0.48	0.57	0.45	0.52	0.52
Top 80%	0.63**	0.63**	0.60*	0.63**	0.57	0.48	0.57	0.45	0.49	0.53
Top 90%	0.62**	0.64**	0.58	0.63**	0.59*	0.47	0.58	0.47	0.57	0.55
All	0.62*	0.63**	0.59*	0.66**	0.60*	0.47	0.57	0.48	0.57	0.57
Median	0.62*	0.65**	0.55	0.63**	0.57	0.45	0.59	0.51	0.56	0.58
Dist	0.62*	0.65**	0.55	0.63**	0.57	0.45	0.59	0.51	0.56	0.58
			Panel E 1960-2001							
Best	0.56*	0.54	0.55*	0.53*	0.53					
Top 1%	0.56*	0.55*	0.55*.	0.53*	0.52					
Top 5%	0.55*	0.55*	0.55*	0.54*	0.54**					
Top 10%	0.55*	0.56*	0.55*	0.53	0.55**					
Top 20%	0.55*	0.54	0.54*	0.56**	0.53					
Top 30%	0.56**	0.54*	0.55**	0.55*	0.53*					
Top 40%	0.57**	0.55*	0.55**	0.56**	0.53					
Top 50%	0.57**	0.55*	0.55**	0.56**	0.54					
Top 60%	0.57**	0.55*	0.54*	0.55*	0.54					
Top 70%	0.57**	0.57**	0.54*	0.56**	0.54					
Top 80%	0.57**	0.57**	0.53	0.56**	0.54					
Top 90%	0.56**	0.57**	0.53	0.57**	0.55*					
All	0.56**	0.56*	0.54*	0.58**	0.56**					
Median	0.55**	0.57**	0.53	0.57**	0.56*					
Dist	0.55**	0.57**	0.53	0.57**	0.56*					

Table 3.3: Diebold-Mariano Test of Equal Predictive Ability between Thin and Thick Modelling Excess Return Forecasts

Each Panel reports the proportion of times that in a given sample the sign of realized excess returns is correctly predicted by the sign of alternative thin and thick modelling one step ahead forecasts generated by five different estimation strategies. '**','*' indicate significant evidence of market timing at the 5%, and 10% levels, respectively. Top x% is the combination of the trimmed mean of the best x% forecasting models. REC, ROLL, OW denote recursive estimation, rolling estimation with fixed window length, optimal estimation window, respectively. The numbers in square brackets show the number of focal variables considered. [1] is just the constant, while [4] denotes the following set of regressors: constant, log of the price-earning ratio, yield-to maturity on long term bonds, yield on 12-month Treasury Bills

	REC [1]	ROLL [1]	REC [4]	OW [1]	OW [4]	REC [1]	ROLL [1]	REC [4]	OW [1]	OW [4]
	Panel A: 1960-1970					Panel B 1970-1980				
Top 1%	-1.19	-0.29	0.04	-1.88	0.01	-0.73	-1.66	0.36	-0.92	0.19
Top 5%	-0.82	-1.18	-0.51	-2.66*	-0.92	-0.20	-1.97	0.35	-3.05**	-1.19
Top 10%	-1.08	-1.65	-0.84	-2.49	-1.08	-0.24	-2.58	0.48	-3.37**	-1.63
Top 20%	-0.84	-2.00	-1.13	-2.57*	-1.27	-0.65	-3.30*	0.25	-2.93*	-1.70
Top 30%	-1.02	-2.23*	-1.48	-2.66	-1.37	-0.57	-3.73**	-0.27	-2.80**	-1.69
Top 40%	-1.04	-2.36*	-1.65	-2.67*	-1.48	-0.83	-3.80**	-0.07	-2.79*	-1.62
Top 50%	-1.13	-2.41*	-1.65	-2.65*	-1.54	-0.98	-3.87**	0.40	-2.79*	-1.75
Top 60%	-1.18	-2.46*	-1.61	-2.62	-1.58	-0.98	-3.81**	0.49	-2.78*	-1.81
Top 70%	-1.13	-2.51**	-1.52	-2.58*	-1.65	-1.08	-3.71**	0.51	-2.78*	-1.87
Top 80%	-1.02	-2.51*	-1.44	-2.55*	-1.67	-1.06	-3.66**	0.57	-2.78**	-1.84
Top 90%	-0.96	-2.47*	-1.39	-2.54*	-1.73	-1.00	-3.59**	0.60	-2.72*	-1.79
All	-0.97	-2.45*	-1.38	-2.58*	-1.72	-0.88	-3.52**	0.69	-2.66**	-1.73
	Panel C 1980-1990					Panel D 1990-2000				
Top 1%	1.10	-1.04	0.50	-0.60	-0.61	0.33	0.45	0.92	-0.02	-2.31
Top 5%	-0.76	-2.20*	0.53	-2.21*	-0.17	0.84	-1.26	-1.22	-1.21	-2.88*
Top 10%	-1.30	-2.91**	-0.26	-2.28*	-0.26	1.51	-2.10	-0.86	-1.70	-3.29*
Top 20%	-1.28	-3.32**	-0.96	-2.24*	-0.49	1.80	-2.75**	-0.87	-2.04*	-3.20**
Top 30%	-1.42	-3.47*	-0.69	-2.21*	-0.93	1.84	-2.93*	-1.26	-2.15*	-3.70**
Top 40%	-1.34	-3.93**	-0.61	-2.27*	-1.57	1.67	-3.03**	-1.40	-2.32*	-3.98**
Top 50%	-1.33	-4.11**	-0.50	-2.32*	-2.10*	1.44	-3.07*	-1.51	-2.40*	-4.00**
Top 60%	-1.32	-4.25**	-0.45	-2.30*	-2.29*	1.11	-3.12**	-1.57	-2.40*	-4.02**
Top 70%	-1.30	-4.26**	-0.35	-2.31	-2.44*	0.88	-3.21**	-1.61	-2.39*	-4.02**
Top 80%	-1.29	-4.18**	-0.32	-2.29*	-2.38*	0.62	-3.28**	-1.62	-2.41*	-3.99**
Top 90%	-1.21	-4.06**	-0.37	-2.29	-2.38*	0.26	-3.29**	-1.63	-2.42*	-3.95**
All	-1.16	-3.96**	-0.43	-2.26*	-2.39*	-0.22	-3.29**	-1.61	-2.40*	-3.92**
	Panel E 1960-2001									
Top 1%	-1.67	-1.42	-0.26	0.29	-0.29					
Top 5%	-5.21**	-2.21*	-1.86	-2.36*	-0.24					
Top 10%	-5.34**	-2.98**	-1.72	-2.17*	-0.37					
Top 20%	-6.21**	-3.58**	-3.13**	-1.93	-0.56					
Top 30%	-6.36**	-3.79**	-3.41**	-1.75	-0.66					
Top 40%	-6.57**	-4.08**	-3.13**	-1.75	-0.90					
Top 50%	-6.45**	-4.09**	-2.95**	-1.77	-1.14					
Top 60%	-6.23**	-3.88**	-3.06**	-1.73	-1.29					
Top 70%	-6.01**	-3.62**	-3.00**	-1.70	-1.47					
Top 80%	-5.79**	-3.42**	-2.93**	-1.64	-1.44					
Top 90%	-5.56**	-3.21**	-2.85**	-1.51	-1.35					
All	-5.09**	-3.05*	-2.81*	-1.37	-1.30					

Table 3.4: White Bootstrap Reality Check

The statistics reported in this table are computed across eleven thick modelling based forecasts, and five estimation strategies (recursive, rolling, and rolling with optimal chosen window estimation with the constant as the only focal variable; recursive and rolling estimation with optimal chosen window and four focal variables. The table reports p-values for the null that thin modelling based forecasts outperform the available thick modelling based forecasts.

	Min	10%	25%	Median	75%	90%	Max
RC p-value	0.000	0.000	0.000	0.004	0.038	0.156	0.429

Figure 3.1: Excess Return on S&P 500

The figure reports monthly excess return on the S&P 500 Index. The sample period is 1955-2001.

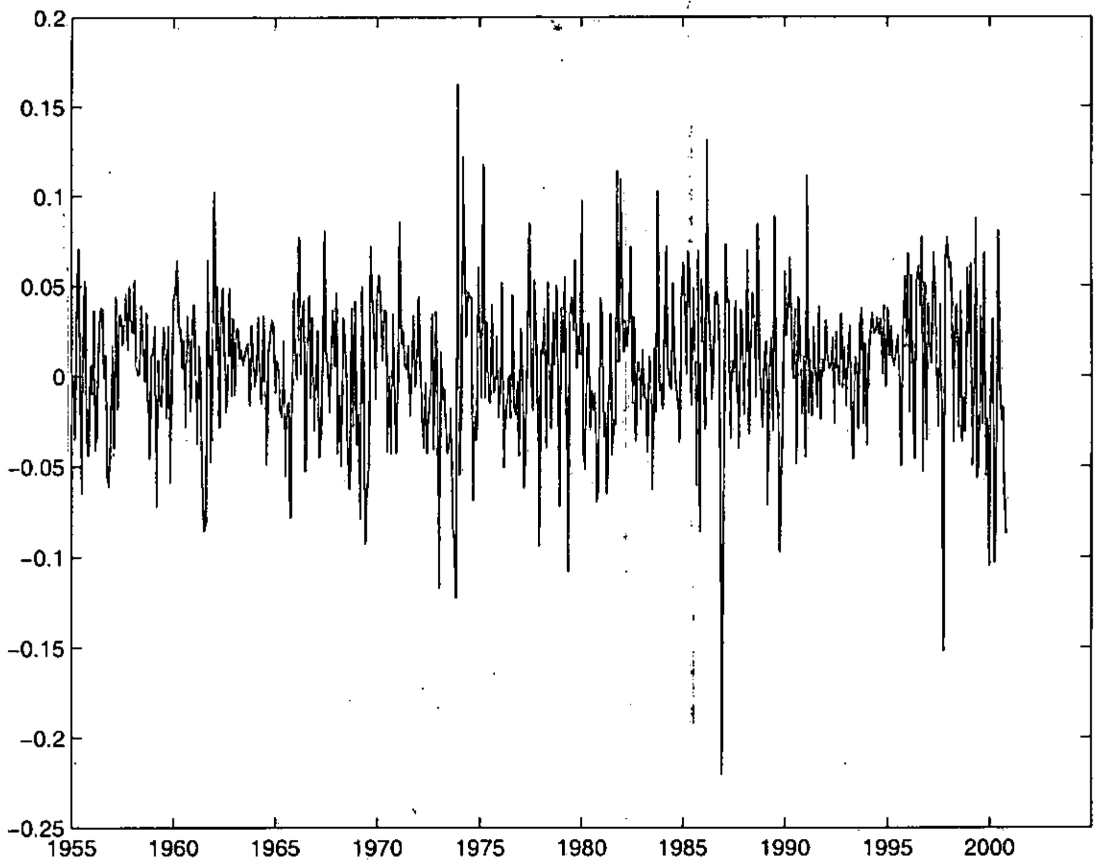


Figure 3.2: Adjusted R^2

The Figure reports the panel of the time-varying adjusted R^2 for the 2^k available models estimated recursively. The first observation refers to the smallest sample (1954.1-1959.12), the last observation refers to the full sample (1954.1-2001.8). The vertical line in 1992.12 shows the results for the Pesaran and Timmermann sample.

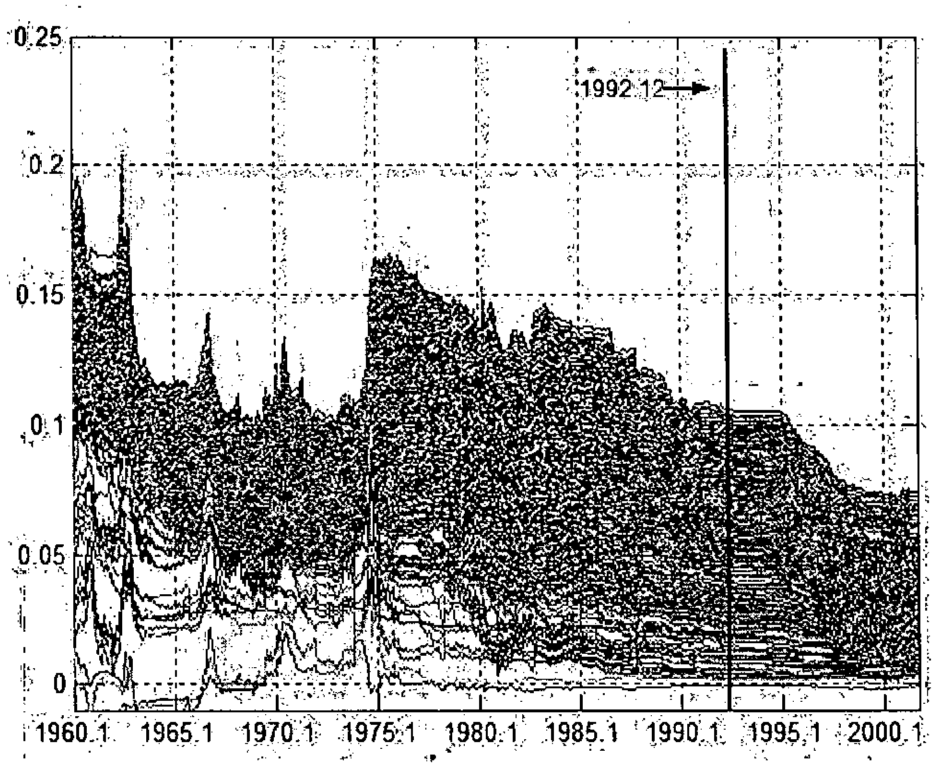


Figure 3.3: Cumulative wealth obtained from all possible portfolio allocations

The figure reports the Cumulative wealth obtained from all possible portfolio allocations. Allocations are associated to models ranked according to their adjusted R^2. The thick line pins down the final wealth delivered by the buy-and-hold strategy.

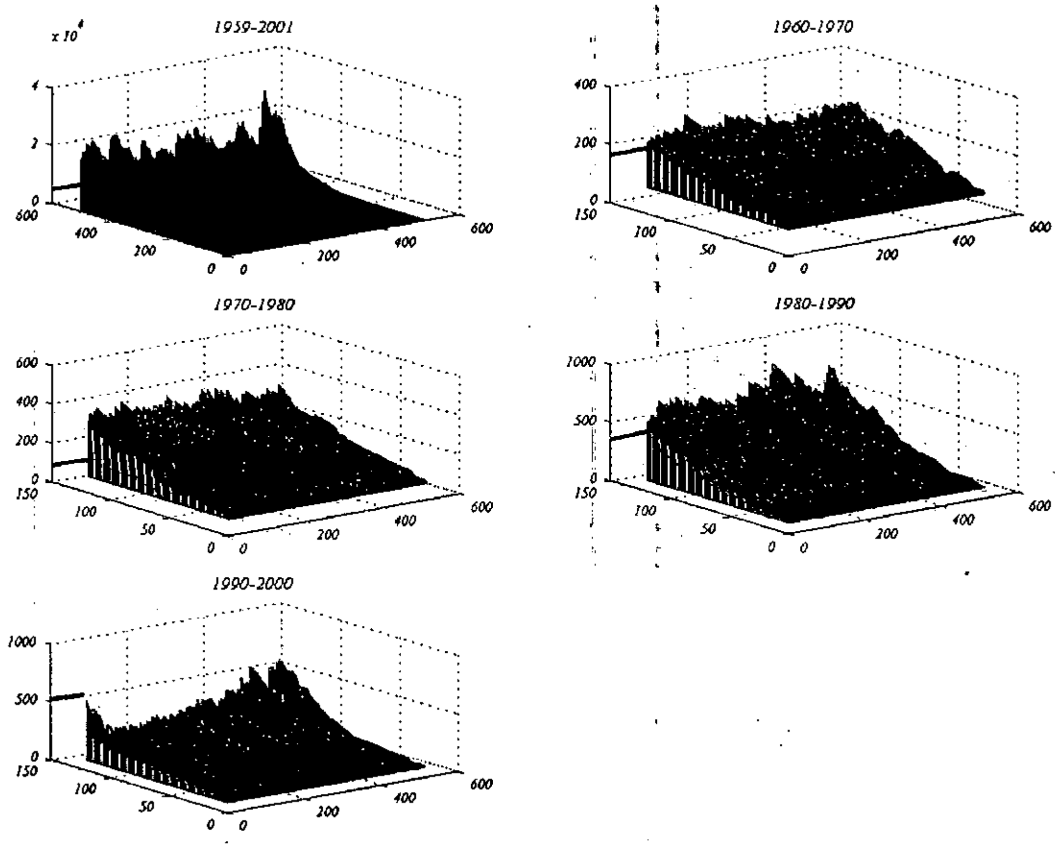


Figure 3.4: End of Period Wealth generated by asset allocation based on thin and thick modelling

Forecasts for excess returns are based on recursive estimation with one focal variable. On the horizontal axis we indicate the thickness of our approach in terms of the percentage of models (ranked by their within sample performance) used in the construction of the different trading rules. Each panel reports the performance of a buy and hold strategy on S&P500 (Mkt), Distributional Thick Modelling, Meta-Thick-Modelling, and Kernel-Thick-Modelling strategies.

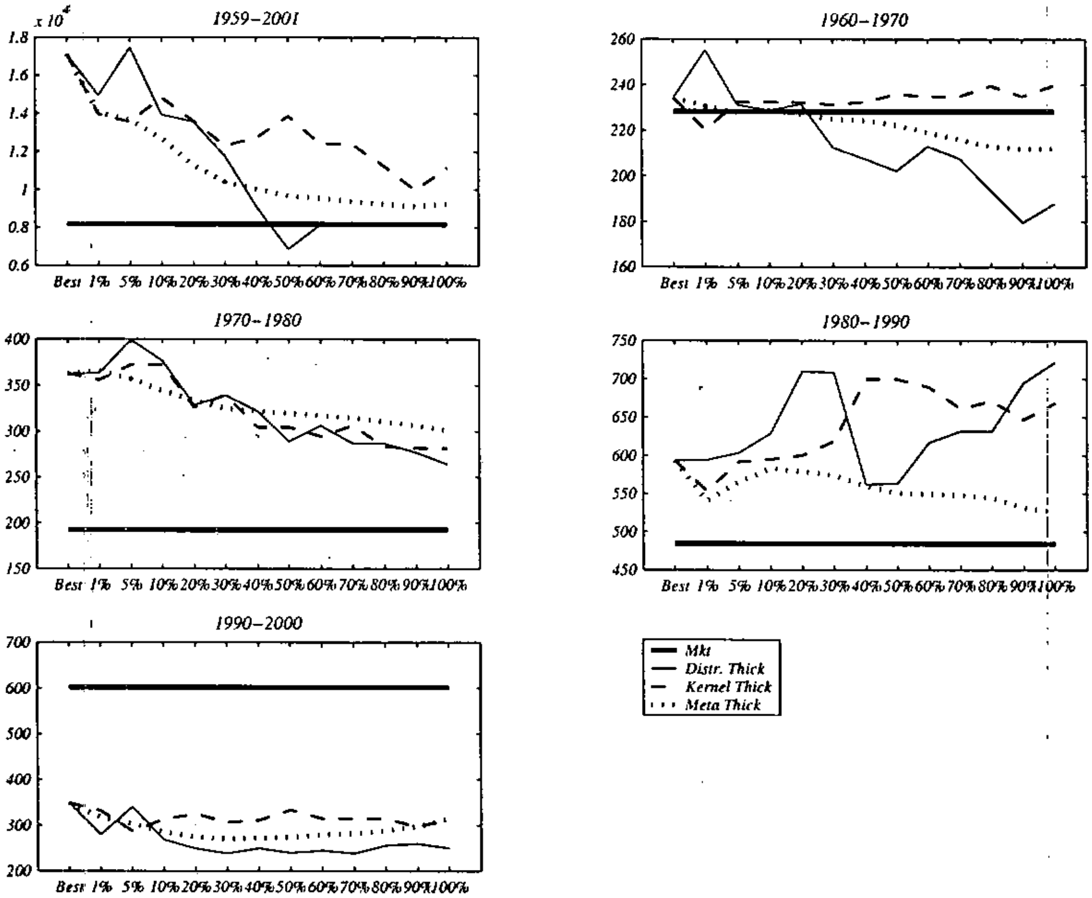
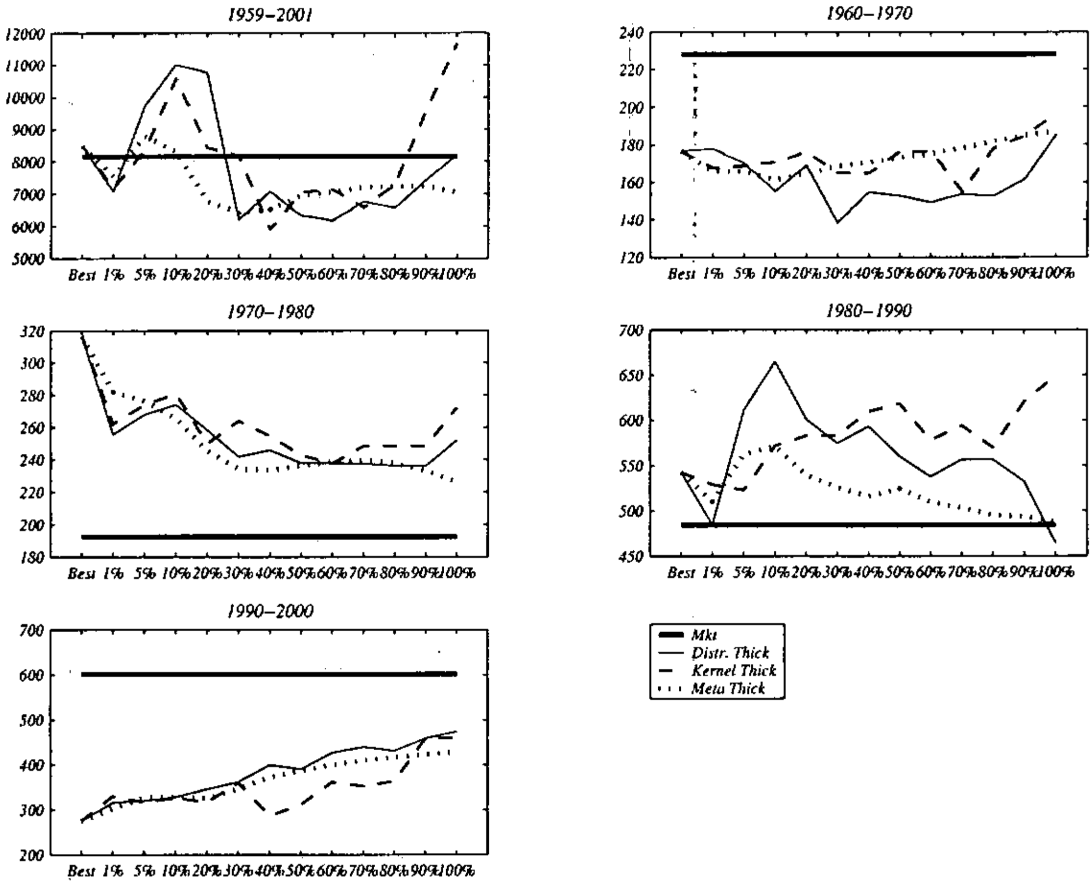


Figure 3.5: End of Period Wealth generated by asset allocation based on thin and thick modelling

Forecasts for excess returns are based on rolling estimation (optimal chosen window) and four focal variables. On the horizontal axis we indicate the thickness of our approach in terms of the percentage of models (ranked by their within sample performance) used in the construction of the different trading rules. Each panel reports the performance of a buy and hold strategy on S&P500 (Mkt), Distributional Thick Modelling, Meta-Thick-Modelling, and Kernel-Thick-Modelling strategies.



Chapter 4

Common Factors in Latin America's Business Cycles

4.1 Introduction

Business[1] cycle volatility can arise from a variety of sources and be exacerbated by distinct economic policy regimes, possibly reflecting slowly-evolving institutional factors (Acemoglu, Johnson, Robinson and Thaicharoe, 2003) and different degrees of financial and trade openness (Kose, Prasad, and Terrones 2005). This suggests that important insights into the phenomenon can be gained from long-run data spanning a variety of policy regimes and institutional settings. Yet there is a striking dearth of systematic work along these lines for most countries outside North America and Western Europe.

[1]This chapter is coauthored with Luis Catão (Research Department, IMF) and Allan Timmermann (UC San Diego). We would like to thank Ayhan Kose, Gian Maria Milesi-Ferretti, Luca Sala, Valerie Ramey, Jeffrey Williamson as well as seminar participants at the Fundação Getulio Vargas in Rio de Janeiro, IDB, IMF, the 2005 LACEA meetings, the 2005 EHES meeting, and at UC San Diego for comments on an earlier draft. We are also grateful to Nathan Balke and Christina Romer for providing us with the US data. The views expressed in this paper are those of the authors and should not be attributed to the International Monetary Fund.

One region that is particularly under-researched is Latin America. This gap is somewhat surprising not only because the region is deemed as highly volatile and the question of what drives such volatility is of interest in its own right; it is also surprising because the region comprises a large set of sovereign nations which have gone through a number of dramatic changes in policy regimes and institutions over a long period of time and relative to other developing countries in Africa and Asia (many of which only became independent nations in recent decades), thus providing a rich context for assessing business cycle theories. Indeed, Latin America is notoriously absent in the well-known historical business cycle studies by Sheffrin (1988) and Backus and Kehoe (1992), and only Argentina is covered in more recent work along similar lines (Basu and Taylor, 1999). Instead, recent research on Latin American business cycles has been either country-specific and covered only short periods of time (e.g. Kydland and Zarazaga, 1997) or focused on specific transmission mechanisms and limited to post-1980 data (Hoffmaister and Roldos, 1997; Neumeyer and Perri, 2005).[2] A corollary of this gap in the literature is the absence of any formal attempt to establish a reference cycle dating for these countries similar to those available for others–such as the United States and the Euro area–on the basis of a variety of coincident and leading indicators (Moore, 1983; Gordon, 1986; Artis, Kontolemis, and Osborn, 1997; Stock and Watson, 1999).

This paper seeks to fill some of this lacuna. Unlike previous work, we go back as far as available macroeconomic data permits and jointly focus on four of the largest Latin American economies - Argentina, Brazil, Chile, and Mexico. Together, these countries have accounted for some 70 percent of the region's GDP over the past half century (Maddison, 2003, pp. 134-140), thus clearly setting the tone for the region's overall macroeconomic performance. At the same time, data availability for this subset of countries permits us to provide a long-run characterization of

[2] A notable exception is Engle and Issler (1993) who use the Beveridge-Nelson trend-cycle decomposition to test for the existence of common trends and common cycles in the real GDP of Argentina, Brazil, and Mexico during 1948-86. They do not provide evidence, however, of what drives the common regional cycle they extract, nor do they look at key variables such as terms of trade or fiscal and monetary shocks that might help explain the observed country-specific movements.

the business cycle in these economies similar to that conducted for advanced countries.

The construction of new indices of economic activity and the identification of volatility sources over such long period allows us to address four main questions. First, how volatile has Latin America been relative to other countries? In particular, has economic activity in Latin America been more or less stable in periods of greater trade and financial integration with the world economy, such as during the pre-1930 gold standard and the post-1980s period? Second, how persistent have macroeconomic fluctuations been in those countries? Since the welfare cost of income fluctuations as well as the burden on stabilization policy rise on volatility and persistence,[3] these are important questions to ask and document. Third, do we observe similar stylized facts as those documented for other economies that feature in the existing business cycle literature? Finally, is there an identifiable regional business cycle?

As discussed further below, a key requirement for answering these questions is to obtain a measure of economic activity that is expected to be reasonably accurate and consistent over such a long period. We provide this by constructing a new index of economic activity for each of the four countries using a dynamic common factor methodology which, to the best to our knowledge, is for the first time applied to build a business cycle index for this set of countries. This methodology is applied to a uniquely large set of macroeconomic variables compiled from a wide range of historical sources. The data span key sectors such as agriculture, manufacturing, mining and cement production, and includes fiscal expenditures and revenues, external variables such as terms of trade, the real exchange rate and import and export volume, as well as a host of financial indicators including interest rates and monetary aggregates. Our index of economic activity is shown to track very closely the existing real GDP data from the full set of national account estimates beginning in the early post-World War II period. Since this index of economic

[3]For theoretical and empirical evidence on the cost of business cycles see, e.g., Van Wincoop (1999) and Pallage and Robe (2003).

activity is constructed as the common factor that underlies a wide set of macroeconomic and sectorial indicators - thus filtering out idiosyncratic components (including possible measurement errors) - it provides a measure that is germane to the concept of the business cycle as defined in the work of Burns and Mitchell (1946) - which still forms the backbone of the widely used NBER reference cycle indicator for the United States.

The paper's main findings are as follows. Over the full sample 1870-2004, the average business cycle volatility in all four countries was considerably higher than in the advanced economies—albeit with important differences over sub-periods. Latin American volatility was relatively high in the pre-1930 era, during the formative years of key national institutions. It then dropped sharply during the four decades following the Great Depression - an apparent pay-off of the inward-looking growth and highly interventionist policy regimes at a time when volatility in advanced countries rose to all-time highs. Cyclical instability in Latin America bounced back again in the 1970s and 1980s - when it was more than twice as high as the advanced country average - before declining sharply more recently. Throughout the period, cyclical persistence has been high, with large shocks giving rise to a striking combination of high cyclical volatility and long business cycle durations relative to advanced country standards.

We also find evidence of a number of regularities highlighted in the existing business cycle literature. In particular, external terms of trade have been strongly procyclical, the trade balance counter-cyclical, and fixed investment has been several times more volatile than output. Using the simple gauge proposed in Kaminsky, Reinhart, and Végh (2004), we also find that fiscal policy has been strongly procyclical in these countries. In contrast with evidence more directly supportive of Phillips curve trade-offs among advanced countries, we find that inflation has been historically counter-cyclical in all four Latin American economies. Compared with the more mixed cross-country evidence in other regions, real wages have also been broadly procyclical. Once again, a contrast with advanced economies lies in the strikingly large volatility of these individual variables.

Concordance indices along the lines of Artis et al (1997) and Harding and Pagan (2002) indicate that business cycles in these economies have been reasonably correlated. Pooling data from all four countries, the common factor methodology that we employ permits the identification of a sizeable common regional factor. Since trade linkages between these economies have been small until very recently and capital account linkages remain so to date, global shocks – notably to key foreign interest rates, real income in advanced countries and commodity terms of trade –emerge as key drivers of this common regional business cycle. This result has salient practical implications that have previously been discussed on the basis of distinct methodologies and far more limited data (Calvo, Leiderman, and Reinhart, 1993; Fernandez-Arias, 1994; Agénor et al., 2000; Neumeyer and Perri, 2005).

The remainder of the paper is divided into five sections as follows. Section 2 lays out the econometric framework and discusses the main estimation issues. Section 3 reports empirical estimates and provides robustness checks of our methodology, while Section 4 presents stylized facts about the business cycle in the Latin American countries. Section 5 concludes. An appendix contains details of the construction and sources of our data series.

4.2 Econometric Framework

The idea that a cross-section of economic variables share a common factor structure has a long tradition in economics, dating back at least to the attempt by Burns and Mitchell (1946) to construct an aggregate measure of economic activity. There are two chief motivations for common factor models. First, economic theory suggests strong linkages between economic activity across different sectors due to common productivity, preference and policy shocks. However, since some of these shocks are unobservable, information about them can only be extracted once one has access to a sufficiently large cross-section of economic variables that are at least in part driven by these shocks.

Hence, a critical requirement that needs to be met in our analysis is the availability of a broad set of variables that bear sufficiently close relation to aggregate business cycle behavior. Natural candidates include capital formation, government revenue and expenditures, sectorial output series, as well as external trade figures and a host of financial variables. The fact that the Latin American economies have historically been highly dependent on global capital markets and demand from outside trading partners suggests that interest rates and cyclical output in advanced countries also be included in the analysis.

The second motivation for using dynamic factor analysis is related to the presence of measurement errors. Activity levels in many sectors are measured with considerable error. Provided that measurement errors are largely idiosyncratic, cross-sectional information can be used to construct more robust common factors that are not similarly sensitive to the impact of such errors. Here one has to make assumptions on the exposure of such observable variables to common shocks in order to identify the underlying driving factors.

Stock and Watson (1989, 2002) and Forni, Hallin, Lippi and Reichlin (2000) have shown that the application of dynamic common factor models to a sufficiently representative set of macroeconomic and sectorial indicators provides superior forecasting performance for a target variable such as real GDP or indeed any broad index of economic activity. This methodology turns out to be particularly useful when some of the constituent series that add up to a target variable (such as monthly GDP) are lacking, or when such series are suspected to be mismeasured (as commonly deemed to be the case for certain service activities). An important requirement is that such measurement errors are sufficiently idiosyncratic or that the cross-section of available time series be sufficiently large and/or representative. This methodology is clearly suitable when interest lies in reconstructing (backcasting) historical measures of the cycle, as discussed below.

4.2.1 Model Specification and Dynamic Factor Estimation

Let $\mathbf{X}_t$ be a vector of de-meaned and standardized time-series observations on N economic variables observed over the sample $t = 1, ..., T$. Assuming that $\mathbf{X}_t$ admits a dynamic factor representation, we can write

$$\mathbf{X}_t = \mathbf{\Lambda}(L)\,\mathbf{f}_t + \mathbf{e}_t$$

$$= [\Lambda_0, ..., \Lambda_s]\begin{bmatrix} \mathbf{f}_t \\ \ldots \\ \mathbf{f}_{t-s} \end{bmatrix} + \mathbf{e}_t = \mathbf{\Lambda}\mathbf{F}_t + \mathbf{e}_t,$$

where $\mathbf{f}_t = (f_{1t}, ..., f_{qt})'$ is a vector of q common dynamic factors, $\mathbf{\Lambda}(L)$ is an $N \times q$ matrix of filters of length s, $\mathbf{e}_t$ is an $N \times 1$ vector of idiosyncratic disturbances, $\mathbf{F}_t = \left(\mathbf{f}_t', ..., \mathbf{f}_{t-s}'\right)$ is an $r \times 1$ vector of stacked factors with $r = q \times (s+1)$. Notice that while q identifies the number of common shocks, the dimension of $\mathbf{F}_t$ depends on the lag structure of the propagation mechanism of those shocks. Similarly, $\mathbf{f}_t$ is the vector of q dynamic factors and $\mathbf{F}_t$ is the vector of r static factors, while $\mathbf{\Lambda}$ contains the factor loadings. We refer to (4.2.1) as the dynamic representation and to (4.2.1) as the static representation.

In practice the factors are typically unobserved and extraction of them from the observables ($\mathbf{X}_t$) requires making identifying econometric assumptions. As is typical in the literature, we assume that the errors $\mathbf{e}_t$, are mutually orthogonal with respect to $\mathbf{f}_t$ although they can be correlated across series and through time. In addition the factors are only identified up to an arbitrary rotation—we explain in the empirical section how we choose a particular rotation using the idea that the factors are only identified indirectly via the factor loadings.

The standard estimation method of dynamic factor models involves maximizing the likelihood function by means of the Kalman filter. This technique has been employed for low-dimensional systems by Stock and Watson (1991). When N is large, non-parametric methods such as static principal components (Stock and Watson (2002)), weighted static principal components (Boivin and Ng, 2003) and dynamic generalized principal components (Forni, Hallin, Lippi and Reichlin, 2000) are available for consistent estimation of the factors in approximate dynamic factor models.

Under the assumed orthogonality between the dynamic factors and the idiosyncratic disturbances, we can consider a spectral density matrix or covariance matrix of the $\mathbf{X}_t$ decomposition and the common component can be approximated by projecting either on the first r static principal components of the covariance matrix (Stock and Watson, 2002) or on the first q dynamic principal components (Forni, Hallin Lippi and Reichlin, 2000), possibly after scaling the data by the covariance matrix (Boivin and Ng (2003)). In this paper we consider both approaches and evaluate the robustness of the results to this choice, since there is no clear-cut evidence on which approach is superior.[4]

In Stock and Watson (2002), a principal component estimator of the factors emerges as the solution to the following least squares problem:

$$\min_{\mathbf{F}_t,\Lambda} T^{-1}\sum_{t=1}^{T}(\mathbf{X}_t-\Lambda\mathbf{F}_t)'(\mathbf{X}_t-\Lambda\mathbf{F}_t)$$

subject to the restriction $\mathbf{\Lambda}'\mathbf{\Lambda}=\mathbf{I}$. The solution to this problem $\{\widehat{\mathbf{\Lambda}},\widehat{\mathbf{F}}_t\}$ takes the form

$$\widehat{\mathbf{\Lambda}}=\nu$$

$$\widehat{\mathbf{F}}_t^{SW}=\nu'\mathbf{X}_t,$$

[4]In their empirical forecasting comparison, D'Agostino and Giannone (2004) find that weighted procedures generally produce better forecasting performance. Similarly, Boivin and Ng (2003) find that weighted principal components improve on the forecasts of the standard principal components methods applied to the static factor model. Stock and Watson (2005) report that forecasts based on factors estimated with static principal components and those estimated with weighted principal components tend to be highly correlated.

where ν is an $r \times 1$ vector of eigenvectors corresponding to the r largest eigenvalues of the variance-covariance matrix of the $\mathbf{X}$-variables, Σ_{xx}. The resulting estimator of the factors, $\widehat{\mathbf{F}}_t^{SW}$, is the first r static principal components of $\mathbf{X}_t$.

In Forni et al (2000) the dynamic structure in the factors is explored by extracting principal components from the frequency domain. Their approach permits efficient aggregation of variables that may be out of phase, with the common component being estimated by projecting the $\mathbf{X}-$variables on present, past and future dynamic principal components. The factors and their loadings are the solution to the following non-linear least squares problem that weights the idiosyncratic errors by their covariance matrix, $\Omega = E[(\mathbf{X}_t - \Lambda \mathbf{F}_t)(\mathbf{X}_t - \Lambda \mathbf{F}_t)']$:

$$\min_{\mathbf{F}_t, \Lambda} T^{-1} \sum_{t=1}^{T} (\mathbf{X}_t - \Lambda \mathbf{F}_t)' \Omega^{-1} (\mathbf{X}_t - \Lambda \mathbf{F}_t),$$

again subject to $\Lambda'\Lambda = \mathbf{I}$. As in Forni et al (2003), we adopt a two-step weighted principal component estimation procedure where Ω is estimated as the difference between the sample covariance matrix, $\widehat{\Sigma}_{xx}$, and the dynamic principal components estimator of the spectral density matrix of the common components.[5]

The resulting estimators of the loadings and common factors are

$$\widehat{\Lambda} = \nu_g$$
$$\widehat{\mathbf{F}}_t^{FHLR} = \nu_g' \mathbf{X}_t = \nu' \widetilde{\mathbf{X}}_t,$$

where ν_g are the generalized eigenvectors associated with the largest generalized eigenvalues of the estimated covariance matrices of common and idiosyncratic components and the resulting estimator

[5]Specifically, let $\mathbf{x}_t$ denote the standardized values of $\mathbf{X}_t$. The estimated spectrum of $\mathbf{x}_t$, $\mathbf{S}_{xx}(\omega)$, is computed at 101 equally spaced ordinates using a Bartlett kernel applied to $p = T^{1/2}$ sample autocovariances. The estimated spectrum of the dynamic factor components, $\mathbf{S}_{ff}(\omega)$, is computed for each of the 101 frequencies using q dynamic principal components of $\mathbf{S}_{xx}(\omega)$. The estimated value of Ω is computed as $\Omega = \Sigma_{xx} - \Sigma_{ff}$, where Σ_{xx} is the sample second moment matrix of x and Σ_{ff} is the inverse fourier transform of $\mathbf{S}_{ff}(\omega)$.

of the factors is the vector consisting of the first r generalized principal components of $\mathbf{X}_t$. This can be seen as the first r static principal components of the transformed data $\widetilde{\mathbf{X}}_t = \left(\hat{\Omega}\right)^{-1/2} \mathbf{X}_t$.

An important requirement when applying these estimators is that all the variables entering the dynamic common factor specification are stationary. With the exception of the inflation rate, real interest rates, and the ratios of export to import value which are stationary by construction, we employ two alternative approaches to ensure stationarity. One is the standard Hodrick-Prescott filter, with a smoothing factor set to 100, as is common practice with annual data (e.g.. Backus and Kehoe, 1992; Kose and Reizman, 2001). The second approach to detrending considered here is the symmetric moving average band-pass filter advanced by Baxter and King (1999). Following common practice with annual data, we set the size of the symmetric moving average parameter to three but use a larger-than-usual bandwidth ranging from 2 to 20 years so as to avoid filtering out the longer (12-20 year) pre-war cycles first documented by Kuznets (1956) for the United States and found to be present in several advanced countries (Solomou, 1987). As shown below, both detrending methods yield very similar results.

4.2.2 Backcasting Historical Activity Measures with Dynamic Factor Models

The common factors derived above, $\widehat{\mathbf{F}}_t^{SW}$ or $\widehat{\mathbf{F}}_t^{FHLR}$, are of interest in their own right since they provide broad-based measures of economic activity. However, often particular interest lies in analyzing a particular time-series such as real GDP over long periods of time. However, data on this variable may only be available over a more recent sample. and, even when available, the series may be subject to considerable measurement error.

The common factor approach is ideally suited to handle these problems provided that the variable of interest lends itself to a similar dynamic factor representation as assumed above. Letting the real GDP cycle be represented by the variable y_t, and under the assumption that y_t is driven by the common factors $\mathbf{f}_t = (f_{1t}, ..., f_{qt})'$ derived above, we have

$$y_t = c + \mathbf{b}(L)\,\mathbf{f}_t + \epsilon_t$$

Our interest lies in backcasting values to create a new historical time-series of cyclical aggregate output so we estimate the following backcasting equation using contemporaneous factor values:

$$y_t = \alpha + \beta'\widehat{\mathbf{F}}_t + \epsilon_t.$$

When data of sufficient quality on y_t is only available over a much shorter (recent) sample than data on the variables used to construct estimates of the factors, under the maintained model , we can estimate the parameters $\theta = \{\alpha, \beta\}$ over a (recent) sample period for which quality data is available on output, y. We can then backcast cyclical output over the longer sample for which estimates of the factors are available. In the following we explain details of how we set up the data and how we deal with estimation issues pertaining to the number of factors and parameter instability.

4.3 New Business Cycle Indices for Latin America

A full set of national income account data for Argentina, Brazil, Chile and Mexico is only available from the mid-1930s (Argentina) or starting at some point in the 1940s for the other three countries.[6] Previous researchers have tried to overcome this limitation by constructing proxy measures of economic activity for the earlier period. The quality of these constructs is, however, very uneven due to the lack and/or the very poor quality of output data for broad sectors of the economy. In the case of Argentina and Brazil, for instance, official output data in agriculture, manufacturing,

[6]Even for Argentina, full-fledged information underpinning national account estimates is not available before 1950 (see Banco Central de Argentina, 1976). In the case of Mexico, a GDP series constructed solely on the basis of sectoral output information—and not based on expenditure and income data—has been reported by Banco de Mexico since 1921.

construction, and services only become available from 1900 onwards and, even then, with serious gaps particularly in the case of Brazil (c.f. Haddad, 1978). With regard to Chile and Mexico, sectorial output data stretching back to the 19th century are more readily available but, again, often spanning a small subset of the universe of firms and of questionable quality (see the Appendix). Insofar as previous researchers tried to derive an aggregate measure of economic activity from averages of these production data (resorting to linear interpolation to fill gaps in some discontinuous annual series), the resulting indices are bound to be highly inaccurate. While two other attempts have been made to overcome these problems, they have clear drawbacks. One is that of backcasting Argentine GDP based on a handful of production and trade variables by means of linear OLS regressions (della Paolera, 1989, p.189); the other is the use of static common components to backcast 19th century Brazilian GDP on the basis of foreign trade data (Contador and Haddad, 1975).[7] Despite this very limited variable span, the latter series has been (misleadingly) compiled by Maddison (1995, 2003) and Mitchell (2003) as a reliable indicator of pre-war Brazilian GDP.

Our paper addresses these data limitations by substantially broadening the number of variables from which one can derive valuable information on the pace of aggregate economic activity. We take into account not only production or foreign trade variables, but also monetary and financial indicators that economic theory suggests should be correlated with the business cycle. As discussed in the Appendix, the data was obtained from an extensive compilation of both primary and secondary data sources. In some cases this resulted in entirely new series being created; once combined with their counterparts from the later 20th century, these series span the entire 1870-2004 period. Still, as one might expect from country specific idiosyncrasies in data collection (especially before the standardization of national account and balance of payments methodologies), the availability of macroeconomic and financial indicators varies somewhat across countries. For example, for Mexico very few variables were measured prior to 1877, so our business cycle index

[7] A cruder attempt of reconstructing 19th century Brazilian GDP can be found in Goldsmith (1986), who derives a GDP growth series based on an unweighted average of government expenditure, revenues, wages, exports and imports.

for that country only starts in 1878. Likewise, it proved impossible to obtain any meaningful series for manufacturing and agriculture output in Brazil before 1900, although we were successful in filling the gap regarding domestic cement consumption (a proxy for construction activity) as well as output in the transportation and communication sectors. A similar gap was filled for pre-1900 Argentina which also benefitted from the use of a new proxy indicator of manufacturing activity starting in 1875 and recently compiled by della Paolera and Taylor (2003).

Overall, we were able to put together a panel of between 20 to 25 time series per country which, as shown below, appears to provide an excellent gauge of the respective national business cycles. The Appendix provides a detailed discussion of measurement issues underlying the various series and the respective data sources. Since we are concerned with real economic aggregates, all variables are measured in real terms deflated by the consumer price index or by the investment or GDP deflators as appropriate - the obvious exceptions being inflation, the ratios of exports to imports, and country spreads (as measured by the difference between the yield on a sovereign foreign-currency denominated bond and the respective UK or US yields).[8]

4.3.1 Empirical Results

Factors extracted from a dataset comprising information on a variety of variables are not typically straightforward to interpret. Nevertheless, the estimated factor loadings do offer important clues in this respect. While factors are only identified up to an arbitrary rotation, it becomes clear from the individual factor loadings that the first factor bears a strong positive correlation with the GDP cycle during periods for which actual GDP data is available.

[8] In the case of interest rates, we employ the commonly used method of defining the real interest rate as the difference between the nominal interest rate and current inflation. Since all interest rate series used in the estimation refer to short-term instruments, discrepancies arising from possible mismatches between current and expected inflation are less critical than in the case of long bonds. Yet, we also checked the robustness of our results to the use of the US 10-year bond yield instead of the 3-month US treasury bill and found that this did not have any effect on inferences made.

Table 1 shows the estimated factor loadings for the first two factors extracted using the Stock and Watson procedure and the HP-detrending since the results using other methods yield very similar estimates, as shown below. We report only the first two factors since the addition of further factors only contributes marginally to the total variance of the panel with the exception of one country (Brazil) for which the third factor is important (more on this below). The first factor (labeled F1) can be interpreted as a broad measure of cyclical activity since it loads positively on indicators that are well-known to be procylical such as sectorial output, fixed capital formation, import quantum and real money, all measured in deviations from their respective long-term trends. The interpretation of the second factor (F2) is less clear-cut. For Argentina, Brazil and Chile, this factor assigns large loadings to money, the domestic interest rate and the real exchange rate (also entered in deviations from trend). Thus, it can be broadly interpreted as an index of monetary conditions. In the case of Mexico, the largest loadings are observed on the variables capturing external linkages such as the term of trade, the real exchange rate or import volume. This is suggestive of an important difference between the economies, possibly indicating that Mexico's linkage to the US economy is of special relevance - a conjecture that is corroborated by further evidence presented below.

Figure 1 plots the two SW factors for each of the countries using the HP detrending as reported in Table 1. For comparison, we also plot the same factors using band-pass filter detrending. Since the two approaches yield very similar results and given that the HP-detrending has been more extensively used in related studies (Backus and Kehoe, 1992; Kydland and Zarazaga, 1997; Kose and Reizman, 2001; Neumeyer and Perri, 2005), we maintain this detrending method through the remainder of the paper.

While the factors are of interest in their own right, ultimately our interest lies in reconstructing a measure of cyclical activity in the Latin American economies. To this end, Table 2 reports the $\bar{R}$−value of regressions of de-trended actual GDP on the factors across a range of factor model specifications. The results cover the period 1950-2004, when full national account estimates are available for all four countries. As with the bulk of the series entering the alternative factor specifications, actual GDP is also expressed in deviations from an HP trend - a widely used measure of the output gap. Correlations in Table 2 thus gauge the extent to which the various factor models span the real GDP cycle. To indicate the sensitivity of the results to the adopted econometric methodologies, we present results both for the all-regressor approach—which maximizes the R^2 by projecting cyclical GDP on all variables—and a range of alternative factor approaches such as the Stock and Watson approach using between one and four common factors and the Forni et al (2000) approach estimated with up to two dynamic factors and up to four static factors. As we shall see later, the high in-sample fit of the all-regressor, "kitchen sink" approach comes at the cost of overfitting the data and producing poor out-of-sample performance.

The linear projections of the GDP cycle on the various factors yield a tight fit for Argentina, with $\bar{R}^2$-values varying from 0.89 for the all-regressor approach to around 0.80 for the two factor approaches. Correlations are also generally high for Chile and Mexico, with 75-85 percent of the variance of the real GDP cycle explained by the first two factors. The fit for Brazil is relatively worse overall, but by including the series on agricultural and manufacturing output (both of which are only available from 1900 onwards), one can raise it to above 70 percent using the SW and FHLR approaches, c.f. panel B of Table 2.

Further evidence that the various approaches tell a similar story can be gleaned from Figure 2, which show the backcast estimates of cyclical GDP in the four economies. In each case the upper panel plots our estimates and (where available) other estimates of cyclical GDP over the period 1870-1950, while the bottom panel shows the corresponding values for the remaining part of the sample. The close proximity between the fitted and actual values for the post-war period is clear from these plots—visual differences only emerge during rare and extremely large spikes such as in Brazil in 1961-62 and 1980. Overall, however, it is plain that: (i) the estimated values closely track actual cyclical GDP whenever this is available; (ii) the various factor approaches generate quite similar estimates of the cycle and (iii) factor estimates often differ substantially from estimates based on the all-regressor least squares approach which is the furthest away from "actual" values, as judged by the observations from better quality estimates of actuals out-of-sample (as for Argentina over 1935-49, Chile during 1940-49, and Mexico 1925-49).[9] This strongly cautions against the use of a kitchen sink approach by researchers in the reconstruction of earlier GDP data.

4.3.2 Robustness Analysis

It is important to investigate whether our estimates are robust to potential instability of the factor loadings and to changes in the factor specification. The common factor projections were built under the assumption that factor loadings remain constant over time. In the same way that out-of-sample forecasts rely on an implicit assumption that certain relationships between predictor variables and the target variable remain constant over the forecasting period, backcasting economic activity measures without this assumption is infeasible.

An advantage of our approach is that the use of common factors can be expected to be reasonably robust against the structural instability that plagues low-dimensional forecasting regressions. Stock and Watson (2002) provide both theoretical arguments and empirical evidence that principal component factor estimates are consistent even in the presence of temporal instability in the individual time-series used to construct the factors provided that this instability averages out in the construction of the common factors. This occurs if the instability is sufficiently idiosyncratic to the various series.

[9]The gaps between the "other" and our backcast estimates for Brazil before 1930 is not surprising since the reported pre-1930 estimate by Haddad (1978) is constructed with very incomplete sectoral data, giving excessive weight to highly cyclical subsectors (like crop production) at the expense of less cyclical ones (like services), in addition to relying extensively on interpolation to fill some gaps. Given the reasonably tight fit between our index and the official GDP data after 1950, we suspect that those differences reflect the inaccuracies of the Haddad index rather than of our index, which relies on information across a wider spectrum of variables.

To evaluate the robustness of our results for the backcasted GDP values, Figure 3 plots the minimum and maximum value across different specifications of the backcasting equation. In particular, we consider:

- Two different estimation samples for the backcasting equation, 1915-2004 and 1950-2004 (GDP data for Chile and Mexico are available only after 1940 so for these countries the backcasting equation is estimated only over the sample 1950-2004.)
- Six different factor specifications: SW(r), $r = 2,3$ (where r is the number of static factors), FHLR(q, r), where the first argument, q, is the number of dynamic factors while the second, r, is the number of static factors. We set $(q, r) = (2, 1), (3, 1), (2, 2)$ and $(2, 3)$.
- Different samples for factor estimation, where a new sample is adopted if new time series become available (Argentina: 1870-2004, 1875-2004, 1900-2004; Brazil: 1870-2004, 1900-2004; Chile: 1870-2004; Mexico: 1878-2004).
- Two different panels of data, one including the external variables while the other excludes these.

The sensitivity analysis produces 72 specifications for Argentina, 36 for Brazil and 12 for Chile and Mexico. With the exceptions of Brazil in 1890-91, 1986 and 1989, Chile in 1929-32 and Mexico in 1916, the range is very narrow; and even for those outlier observations, all estimates point in the same direction. As it turns out, all indications are that little has changed over time. This congruence would be unlikely to hold if the factor loadings were subject to structural breaks or considerable instability.

In addition, we have also checked for the stability of coefficients in the regression of the factors on the cyclical component of real GDP (equation (5)). This was done by re-estimating the regression for the period 1961-2004 (instead of 1950-2004) and recursively rolling back the estimation to the last point for which reasonably reliable data on real GDP exists.[10] The results plotted in Figure 4 show that the backcasting equation coefficients are reasonably stable over the 1930-60 period (1940-60 for Chile); only in the case of Mexico between 1921 and 1925 is there evidence of some instability. But since the real GDP figures used to compute the recursion over the period in the early post-revolutionary period for Mexico are likely to be marred by measurement problems before the Banco de Mexico centralized the compilation of macroeconomic data in 1925, this should not be surprising.

Overall, the results above make a simple but important point. Even when the common factors are extracted from a dataset containing a limited number of series on output growth, they track the real GDP cycle well. This may not be overly surprising since we selected variables that economic theory suggests should be closely related to cyclical activity. Yet, this evidence underscores the robustness of backcasting inferences on the aggregate output cycle once they are based on a sensible combination of fiscal, financial, sectorial and external variables.

4.3.3 Gains from Using Extended Data Set

Although our results do not appear to be sensitive to the particular choice of factor estimation methodology, number of factors or sample period used to estimate the factor loadings or projection coefficients, one might ask what the value-added is of using as wide a set of variables as that adopted here when constructing the common factors. To answer this, we compare in Figure 5 plots of the first common factor constructed using our extensive data set on sectoral output, financial, fiscal and trade variables against that using only sectoral output variables. Common factors based exclusively on sectoral output data are far smoother than those based on the wider set of variables.

[10]While, as discussed in the text, pre-war data on GDP for all the countries are considerably less reliable than post-war official data, we thought it would still be worthwhile to compare the stability of the backcasting regression coefficients against some of the existing pre-war data as a further robustness check. Given that data for Argentina, Brazil and Mexico from the 1920 onwards appear to be of much better quality (albeit still relying on partial production data) than pre-1920 data, we extended this recursive stability test to 1920 using this data.

This shows up in a failure of the more narrowly constructed common factors in fully accounting for the depth of the crises in Argentina in 1918 and 1990, in Brazil and Mexico following World War I and in Chile following the Great Depression. In addition many of the smaller peaks and troughs - such as the cycle around 1900 in Argentina - are entirely missed by the common factor based on sectoral output information.

This limitation of the smaller set is not exclusive to Argentina for which we have only two sectoral variables going back to 1870. Adding industrial output for Argentina (a series that becomes available from 1875) does not overturn this conclusion. Significant gaps also arise for Brazil, Chile, and even Mexico which has a wider sectoral output data coverage back to the 1870s. Discrepancies between the two series are not exclusive to the pre-war period, and hence do not seem entirely attributable to the poorer quality of earlier data; large gaps emerge, for instance, for post-1960 Brazil.

These plots vividly demonstrate the importance to the construction of broad measures of economic activity of using a wide and varied set of economic variables representing not just a few sectoral output series. In other words, fiscal, financial and external trade variables play an important role in filling the gap.

4.3.4 Actual vs. Backcasted series: A Test based on US data

Skeptical readers might object that we have not, so far, provided any direct evidence that our approach works well in terms of backcasting the cycle. This is true in the sense that the same absence of reliable and broad based historical data on output in Latin America that motivated our analysis also makes it impossible for us to compare our fitted values against realized observations.

To address this concern, we used our approach on the US pre-war cycle. Since the US has high quality real GDP estimates going back to 1870, we can directly compare model predictions and "actual" values (or, to be precise, resonably accurate estimates of the "actual" value). For this test to be informative for our analysis, it is important that we use a set of US variables similar to those used for the four Latin American countries, even though much greater data availability for the US would have allowed us to include many more variables in the estimation of the factor model. With this consideration in mind, we stack the deck against our approach by backcasting the US pre-war cycle based on an even smaller set of variables (18 in total, as described in the Appendix) than those for the four Latin American countries.

Table 3 reports the R^2-values of the various estimation methods.[11] The fit of the various factor models over the 1950-2004 period is very good, with $\bar{R}^2$--values around 0.80. The last column also indicates that our backcasted cycle closely tracks Balke and Gordon's (1989) revised estimate of US real GDP, which we detrend by an HP filter and plot in Figure 6 together with our estimates.[12] Clearly the fit is not perfect - with the largest discrepancy emerging during the World War II boom; yet, the overwhelming majority of cyclical turning points is consistently picked up by both indices.[13]

This exercise indicates that application of our backcasting methodology to a sufficiently representative set of macroeconomic and sectoral variables can yield a very close proxy of actual cyclical fluctuations in US real GDP. To the extent that the in-sample fit for the post-war period is even higher for some of the Latin American countries shown above—and recalling that the span of variables is larger—this suggests that our approach is very likely to be picking up turning points and cyclical variations in these economies quite accurately.

[11] The loadings of the fitted model over 1870-2004 are not reported to conserve on space but are available from the authors upon request.

[12] As shown in Table 3, using Romer's (1989) lower volatility estimate of US pre-1929 GDP, the fit is less tight but only slightly so. For instance, for the SW estimator with 3 factors, the $\bar{R}^2$ using Romer's pre-1929 GDP data is 0.83 as opposed to 0.86 for the Balke and Gordon estimates.

[13] Interestingly, the very high amplitude of the World War II cycle in the Balke and Gordon estimates is not shared by an earlier (and previously widely used) indicator of US GDP by Kuznets (1961). In fact, the Kuznets index (not shown to avoid cluttering but available from the authors upon request) and our index show a positive output gap of around 15 percent in 1944 as opposed to nearly 25 percent in the Balke and Gordon index.

4.3.5 Tracking History

As a final robustness check we ask how well the backcasted series square with qualitative historical evidence on events deemed to be major economic turning points in these countries. Figure 7 relates the two. Starting with pre-war Argentina, the index picks up all economic downturns associated with well-known world events - notably the stock market crashes in Europe and the US in 1873 and the ensuing global economic depression, the 1890 Barings crisis, the 1907 financial panic, the two world wars, and the Great Depression of the 1930s. Likewise, major post-WWII shocks are also conspicuously picked up as turning points in our index, notably the boom and bust in world commodity prices associated with the Korean War in the early 1950s, the oil price shocks of the 1970s, the early 1982-83 debt crisis, as well as the emerging market crises of the 1990s (the 1994-95 "Tequila" crisis and the Asia and Russia crises of 1997-98). A glance at Figure 7 also indicates that such a juxtaposition of cyclical turning points in country indices with major global economic events is broadly corroborated for Brazil, Chile, and Mexico.[14]

In addition, the portrait of history provided by our index is consistent with narrative evidence about the macroeconomic repercussions of key country-specific events. In the case of Brazil, the index picks up the mini downturn associated with the 1888 political unrest (end of Slavery and the republican transition) as well as the subsequent boom (the "Encilhamento") stemming from a liberal monetary reform that brought about an unprecedented boom in domestic credit and asset valuations in 1889-90 (see Trinner, 2000). The Brazil index tracks equally well what is deemed to have been one of Brazil's most protracted recessions which culminated in the country's first sovereign default and the debt rescheduling arrangements under the auspices of the Rothchilds in

[14]In contrast with Argentina, Brazil and Chile were little affected by the 1994-95 Mexican crisis partly due to offsetting domestic developments. In the case of Brazil, a successful stabilization plan in 1994 and renewed political stability set off a domestic demand boom in the following year. In the case of Chile, stronger trade linkages with Asia, low public debt, and a significant improvement in external terms of trade limited the disruptive effects of the Tequila crisis on the domestic economy (see Singh et al., 2005).

1898 (see Fritsch, 1988).[15] As for Chile, our index highlights the upturn of 1879-82 associated with the "War of the Pacific" (against Peru), the downturn around the country's exit from the gold standard in 1898 (Llona Rodriguez, 2000), as well as the severity of the 1929-32 depression in Chile due to plummeting terms of trade (Diaz-Alejandro, 1984). Both in Argentina and Chile as well as (to a lesser extent) Brazil, the index identifies clear turning points around the military coups of the 1960s and 1970s.

Finally, the Mexico index yields a picture of economic fluctuations that is remarkably consistent with that depicted by Mexican historiography starting with the 1879-82 upturn that is typically associated with the onset of the new regime headed by General Porfirio Diaz (Cardenas, 1997). Likewise, the subsequent recession, which takes place in the wake of the US economic slowdown of 1883-84, is clearly depicted; its 1885 trough coincides with the well-documented austerity plan imposed by Diaz's finance minister Manuel Dublan that involved a temporary suspension of payments on domestic public debt (Marichal, 2002). This was followed by an upswing associated with Mexico's renewed access to international capital markets in the wake of the 1886-87 external debt settlement, which was later brought to a halt by a sharp worldwide fall in silver prices (Mexico's main export item) coupled with a severe downturn in the United States and sudden stop in capital flows to emerging markets in the early 1890s (c.f. Catão and Solomou, 2005). Finally, our business cycle index also provides a new measure of the severity of the economic downturn associated with the Mexican Revolution of 1911-20 identifying a trough around 1915-16—these were the years when the revolutionary conflict peaked and chaotic monetary conditions triggered a hyperinflation (Cardenas and Manns, 1987).

[15] Unlike several Latin American countries which defaulted on their external debts (and some also on their domestic debts) more than once throughout the 19th century, Brazil consistently serviced its sovereign debt obligations until this time.

4.4 Stylized Business Cycle Facts

Armed with the business cycle indicators for the four countries, we turn to the task of establishing some stylized facts about the four countries' business cycles. We start with dating the respective turning points. A classic device to this end, which is also consistent with our definition of the business cycle as output deviations from a stochastic or deterministic trend, is the Bry and Boschan (1971) algorithm.[16] It consists of a sequence of procedures starting with the search for extreme values in order to eliminate (near-) permanent jumps in the series associated with outliers, followed by the use of centered moving averages and the search for local maxima or minima within a chosen window length.[17] To permit the identification of both shorter and longer cycles, Panels A and B of Table 4 report results based on two-year and six-year windows, respectively. As expected, the algorithm identifies peaks and troughs that are broadly consistent with a visual inspection of Figure 7. When the narrow window is used, the average duration of the cycle is shorter overall, more so during the post-war era. This finding is consistent with evidence of the shortening business cycle length among advanced countries (see, e.g., Gordon, 1986). Using a longer window, Panel B indicates that the pre-cycle is dominated by the Kuznets or long swings, with similar turning points as those identified in the literature on Anglo-saxon economies (Solomou, 1987). This evidence is further reinforced by spectral density function estimates of the individual country indices, which point to a dominant cyclical length around 14 to 16 years during the 1870-1930 period (a typical Kuznets-swing length), followed by a 10-12 year cycle in post-war data (Table 5).

[16]An alternative dating procedure which also builds on the Bry and Boschan approach has been advanced by Harding and Pagan (2002). Their procedure has been designed for use with quartely data and growth cycles rather than with measures of the output gap. As discussed in Marcellino (2005), measuring the cycle as deviations from trend as we do is a more suitable procedure in contexts where absolute declines in output are not so rare, as in our group of countries over the past century. Conversely, the concept of a growth cycle is more analytically relevant when absolute declines in output are very rare and growth rates are reasonably persistent, as in Western Europe in the early post-war decades. Most of the recent empirical literature on business cycle identification and measurement has focused on the classical cycle defined in terms of deviations from trend (stochastic or deterministic).

[17]See King and Plosser (1984) and Watson (1994) for further details and application to US data, for which the algorithm closely replicates the dating by the NBER's panel of experts.

In sum, both the Bry and Bosham algorithm and the spectral density function estimates point to a reasonably long average cyclical duration in all four countries. The dominant cyclical pattern was generally longer in the pre-1930 era, but even in the post-World War II period, cycles in Latin America were substantially more protracted than in the United States and other advanced countries.

Against this background, Tables 6 and 7 report a set of descriptive statistics that help characterize other stylized facts about Latin America's business cycles from a broad cross-country historical perspective. First, standard deviations corroborate the perception that Latin America has been a more cyclically volatile region than both countries deemed advanced by today's definition as well as countries such as Australia, Canada and Japan that were considered "emerging economies" in the pre-war world. This volatility gap between the two groups has changed over time, however. The four Latin countries were clearly far more volatile in the early globalization period before the 1930s - characterized as it was by free capital mobility and very limited quantitative restrictions on trade. Conversely, there is evidence that the inward growth policies did succeed in fending these countries off global instability in the 1930-70 sub-period, when global volatility generally rose, partly due to the recovery from the 1929-32 depression and war shocks. This appears reflected in the higher standard deviation of the output gap among advanced countries during the period as well as among a group of other developing countries for which pre-war GDP estimates are available (India, Indonesia, Korea, Malaysia, Sri Lanka, South Africa, Taiwan, and Turkey). But as output gap volatility came down in advanced countries in the post-1960s period (notwithstanding two oil shocks and dramatic changes in policy regimes), cyclical volatility in Latin America remained relatively high; only in the post-debt crisis period has Latin American cyclical volatility declined markedly compared to earlier levels. Further, Table 7 shows that this decline in cyclical volatility over the past 15 years or so has not been a preserve of Latin America but is also observed in other regions of the developing world - partly reflecting lower real interest rate and output volatility in the US and other advanced countries (see Table 8). Yet, despite being low relative to its earlier historical record, business cycle volatility in Latin America still remains higher than in advanced countries as well as relative to Asian developing countries. Rolling standard deviations of the output gap in Figure 8 summarize this broad overview of volatility trends in the region by plotting both individual country trends as well as that of the common regional cycle (extracted as discussed below).

Table 8 focuses on key drivers of aggregate business cycles in the four economies, once again broken down by sub-periods. The table clearly highlights some stylized facts that have been stressed in previous studies (Backus and Kehoe, 1992; Mendoza, 1995; Basu and Taylor, 1999; Agénor et al. 2000). First, cyclical volatility in fixed investment is much higher than that of output. Second, and consistent with the findings of Backus and Kehoe (1992) for advanced countries, government spending volatility is higher than output volatility. For all four countries and across all sub-periods, the magnitude of two simple gauges of government-induced volatility—the real government expenditure cycle and the ratio of public expenditure to revenues—is staggering. Coupled with the positive loadings of the real government expenditure variable on the first (pro-cyclical) factor in Table 1—and with all the caveats about some inevitable endogeneity of this or indeed of any measure of the fiscal stance—this provides a *prima facie* case that changes in fiscal stances have been important drivers of the business cycle in these countries. This finding squares well with the post-1960 evidence on strong fiscal procyclicality in these countries provided in Kaminsky, Reinhart and Végh (2004) who use the cyclical component of real government spending as their main gauge.

Third, the volatility of monetary aggregates (expressed in real terms) is smaller than that of the fiscal variables with the exception of Argentina and Brazil over the past two decades and Chile in the 1970s reflecting bouts of high- and hyper-inflation in these countries. Interestingly, however, inflation has been broadly counter-cyclical (see Table 1), in stark contrast with the Phillips-curve trade-off which is usually deemed to hold at least among advanced countries. The counter-cyclical behavior of inflation makes the apparent procyclicality of real wages (see Table 1) consistent both with models based on short-run nominal wage stickiness as well as with real business cycle models which emphasize the dominant role of technology shocks in shifting the labor demand schedule over business cycle frequencies. Finally, terms of trade fluctuations are highly procyclical and, consistent with earlier work (Mendoza, 1995), emerge as an important (and more clearly exogenous) source of output volatility. While this may not be particularly surprising given that all four countries have mainly been primary commodity exporters for much of the period (the manufacturing share of Brazil's and Mexico's exports only became prominent over the past couple of decades), it is still instructive to observe the sheer magnitude of the phenomenon. To the extent that terms of trade volatility has important welfare implications and is usually associated with poorer long-term growth performance (Blattman, Hwang, and Williamson, 2005), this emerges as an important feature of the data.

A final set of stylized facts that we document can be gleaned from a look at Figure 7, which shows that several major business cycle turning points–such as those of the early 1890s, World War I, the early 1930s and the early 1980s–are common to all or most of the four countries. A formal measure of such synchronicity is the concordance index proposed by Harding and Pagan (2002). It consists of a non-parametric measure of the relative frequency at which countries are jointly undergoing an expansion or a contraction phase gauged by a binary indicator. Table 9 reports the respective statistic which ranges from a minimum of zero (no concordance) to unity (perfect concordance). The results indicate that Latin American business cycles have displayed a reasonably high degree of synchronization through the 1870-2004 period. This is especially striking in light of the fact that there has been very little intra-regional trade between these economies until the past fifteen years or so, and that such synchronization did not decline dramatically during the period from the early 1930s to the early 1970s marked by strong trade restrictions and capital controls. These results indicate the presence of a common regional factor superimposed on the distinct country-specific business cycle drivers. To gauge this hypothesis more formally we use the econometric methodology from Section 2 to extract common factors from a pooled data set that brings all four countries' data together.[18] The resulting regional factor jointly loads on the various country specific business cycle indicators. Corroborating the concordance metric of Table 9, the regional factor generates correlation coefficients between 0.6 and 0.75 with the procyclical factor (F1) of the business cycle indices in the four individual countries.[19] This clearly points to a sizeable regional common component. This is consistent with both the importance of external variables in the various countries' individual factor loadings (see Table 1), as well as with a long and distinguished literature on the roles of foreign interest rates, income shocks in advanced countries, and commodity terms of trade in triggering financial crises and, more generally, driving key macroeconomic aggregates in Latin America (Diaz-Alejandro, 1984; Fishlow, 1989; Calvo, Leiderman, and Reinhart, 1993; Fernandez-Arias, 1994; Neumeyer and Perri, 2005).

4.5 Conclusions

This paper has sought to fill some of the lacuna in the international business cycle literature. Taking a century long view of the Latin American business cycle allowed us to characterize a host of stylized facts, compare them with existing evidence for other countries, and identify important differences in business cycle behavior across distinct policy and developmental regimes.

We have shown that Latin America has historically displayed high cyclical volatility compared to advanced country and other relevant benchmarks. Further, this volatility has been time-varying. It was highest during the early globalization era of the late 19th and early 20th century—precisely

[18]The results are essentially the same, irrespective of whether we exclude or include the foreign interest rate and advanced countries' GDP in the panel. The results reported exclude the two external variables, if anything thus stacking the deck against finding sizeable co-movement.

[19]The respective plots and sub-period descriptive statistics are not reported to conserve on space but are available from the authors upon request.

during the formative years of key national institutions—then declined markedly over the four decades since the great depression. Yet, we have also shown that after bouncing back in the wake of the large global shocks of the 1970s and early 1980s (Bretton Woods abandonment, oil price and interest rate shocks), business cycle volatility in Latin America subsequently declined to near historical lows. Since this coincides with greater trade and financial openness, a prima-facie link between business cycle volatility and openness is unwarranted looking at the period as a whole. Lower external volatility in output and interest rates over the past fifteen years or so is certainly a set of factors at play; more stable fiscal and monetary policies is another, which also helps explain the different volatility performances across countries amidst the general downward trend.

The paper's other main finding is that such volatility has been strikingly coupled with high business cycle persistence. Since the welfare costs of business cycles are known to rise on both volatility and persistence, the attendant welfare losses have been non-trivial. While it is beyond the scope of our analysis to probe further into the sources of output persistence in the four countries, there is cross-country evidence suggesting that the role of domestic institutions and their constraints on policy making are key (Acemoglu, Johnson, and Robinson, 2005). In addition, to the extent that external developments have themselves been a main source of such persistent shocks via commodity terms of trade and macroeconomic conditions in advanced countries, this has made the task of stabilization policies all the more difficult throughout the region.

We have also shown that several empirical regularities highlighted in the existing business cycle literature readily apply to the four countries. One set of regularities pertains to the countercyclicality of trade balances and the much higher cyclical amplitude of both fixed investment and real government spending relative to output. Indeed, using the simple yardstick of the co-movement between real government expenditures and the output cycle employed elsewhere (Kaminsky, Reinhart and Végh, 2004), we find that fiscal policy has been procylical in all four countries. A comparison between the volatility of external aggregates such as the output gap in the

G-7 countries and external interest rates, and the volatility of domestic aggregates such as public expenditures and revenues as well as, to a lesser extent, real money indicate, if anything, that domestic amplification mechanisms have played a key role. This suggest that these mechanisms deserves close scrutiny by future research on business cycles in these countries.

Finally, our common factor framework also allowed us to identify a sizeable regional common component in Latin American business cycles. Since trade linkages between these economies have been small well into the 1980s, and as capital market linkages remain so to date, this highlights the role of global factors in driving such a regional common cycle. This evidence is consistent with a long literature on the roles of external factors in both triggering financial crises (e.g. Diaz-Alejandro, 1984; Fishlow, 1988) and driving key macroeconomic aggregates in the region (Calvo, Leiderman and Reinhart, 1993; Fernandez-Arias, 1994; Neumeyer and Perri, 2005), as well as with the findings of more general and recent studies emphasizing the role of international factors in individual countries' business cycles (Kose, Otrok, Whiteman, 2003; Canova 2004). In extending these findings to Latin America based on wider time series evidence, the results presented in this paper further highlight the limited scope that regional risk-sharing has had historically.

Bibliography

Acemoglu Daron, Simon Johnson, James Robinson and Yunyong Thaicharoe, 2003, Institutional Causes, Macroeconomic Symptoms: Volatility, Crises and Growth, Journal of Monetary Economics, 50, 49-123.

Acemoglu Daron, Simon Johnson and James Robinson, 2005, Institutions as the Fundamental Cause of Long-Run Growth, in Aghion and Durlauf (eds.), Handbook of Economic Growth, Forthcoming.

Agénor, Pierre-Richard., John McDermott, and Eswar S Prasad., 2000, Macroeconomic fluctuations in developing countries: Some stylized facts, World Bank Review, 14 (2), 251-86.

Artis, Michael, Zenon G. Kontolemis, and Denise Osborn, 1997, Business Cycles for G7 and European Countries, Journal of Business, 70 (2), 249-79.

Backus, David, and Patrick Kehoe, 1992, International Evidence on the Historical Properties of Business Cycles, American Economic Review 82, 864-88.

Balke, Nathan, and Robert Gordon, 1989, The Estimation of Prewar Gross National Product: Methodology and New Evidence, Journal of Political Economy, 97, 1, 38-92.

Banco Central de la Republica Argentina, 1976, Cuentas Nacionales de la Republica Argentina, Buenos Aires.

Basu, Susanto and Alan M. Taylor, 1999, Business Cycles in International Historical Perspective, Journal of Economic Perspectives 13 (2), 45-68.

Baxter, Marianne, and Robert G. King, 1999, Measuring business cycles: Approximate band-pass filters for economic time series, Review of Economics and Statistics 81:4, 575-59.

Bernanke Ben, Boivin Jean; Eliasz Piotr S. 2005, Measuring the Effects of Monetary Policy:

A Factor-augmented Vector Autoregressive (FAVAR) Approach. Quarterly Journal of Economics, 120, 1, 387-422.

Blattman, Chris, Jason Hwang, and Jeffrey Williamson, 2005, The Impact of the Terms of Trade on Economic Development in the Periphery, 1870-1939: Volatility and Secular Change. Journal of Development Economics, Forthcoming.

Boivin, Jearn and Serena Ng, 2003, Are more data always better for factor analysis? Working Paper No. 9829 (NBER).

Braun, Juan, Matías Braun, Ignacio Briones, José Diaz, Rolf Luders and Gert Wagner, 2000, Economía Chilena 1810–1995: Estadísticas Históricas, Working paper 187, Catholic University of Chile, Instituto de Economía.

Bry, Gerard and Boschan, Charlotte, 1971, Cyclical Analysis of Time Series: Selected Procedures and Computer Program, New York.

Burns, Arthur F. and Wesley C. Mitchell, 1946, Measuring Business Cycles, New York.

Calvo, Guillermo, Leonardo Leiderman and Carmen Reinhart, 1993, Capital Flows and Real Exchange Rate Appreciations in Latin America, IMF Staff Papers, 40, 1, 108-151.

Canova, Fabio, 2005, The transmission of US shocks to Latin America, Journal of Applied Econometrics 10, 229-251.

Cárdenas, Enrique, 1997, A macroeconomic interpretation of nineteenth-century Mexico, in S. Haber (ed.), How Latin America Fell Behind: Essays on the Economic History of Brazil and Mexico, Stanford.

Cárdenas, Enrique, and Carlos Manns, 1987, Inflation and monetary stabilization in Mexico during the revolution, Journal of Development Economics, 27, 275-94.

Catão, Luis, and Solomos Solomou, 2005, Effective Exchange Rates and the Classical Gold Standard Adjustment, American Economic Review, 94, 1259-75.

Contador, Claudio and Claudio Haddad, 1975, Produto Real, Moeda and Preços: A Experiência Brasileira no Período 1861-1970, Revista Brasileira de Estatistica, jul./set. 407-440.

D'Agostino, Antonello, and Domenico Giannone 2004, Comparing alternative predictors based on large panels factor models, mimeo.

della Paolera, Gerardo, 1989, How the Argentine Economy Performed During the International Gold Standard: A Re-examination, PhD dissertation, University of Chicago.

della Paolera, Gerardo, and Alan Taylor (eds.), 2003, A New Economic History of Argentina, Cambridge.

Diaz-Alejandro, Carlos F., 1984, Latin America in the 1930s, in Rosemary Thorp (ed), Latin America in the 1930s : the role of the periphery in world crisis Oxford.

Engle, Robert, F. and Issler, João V., 1993, Common trends and common cycles in Latin Ameria, Revista Brasileira de Economia, 47, 2, 149-76.

Fernandez-Arias, Eduardo, 1994, The new wave of private capital inflows: Push or pull? Journal of Development Economics, 48, 389-418.

Fishlow, Albert, 1989, Conditionality and Willingness to Pay: Some Parallels from the 1890s, in Barry Eichengreen and Peter H. Lindert, The International Debt Crisis in Historical Perspective, Cambridge, Mass.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin, 2000. The Generalized Factor Model: Identification and Estimation. Review of Economics and Statistics 82, 540-554.

Forni, M, M. Hallin, M. Lippi, and L. Reichlin, 2003, The generalized dynamic factor model:

forecasting and one sided estimation, CEPR working paper 3432.

Fritsch, Winston, 1988, External Constraints on Economic Policy in Brazil, 1889-1930, London.

Giannone, D., Reichlin, L., and L. Sala (2004), Monetary policy in real time, forthcoming, NBER Macroeconomics Annual.

Goldsmith, Raymond, 1986, Brasil: 1850-1984: Desenvolvimento Financeiro sob um Seculo de Inflacao, Sao Paulo.

Gordon, Robert (ed.), 1986, The American Business Cycle. Continuity and Change. Chicago.

Haddad, Claudio, 1978, O Crescimento do Produto Real Brasileiro, 1900-47. Rio de Janeiro.

Harding, Don and Adrian Pagan, 2002, Dissecting the cycle: a methodological investigation. Journal of Monetary Economics,49 (2), 365-81.

Hoffmaister, Alexander and Jorge Roldos, 1997, Are Business Cycles Different in Asia and Latin America? IMF working paper 97/9.

Kaminsky, Graciela, Carmen Reinhart, and Carlos Végh, 2004, When it Rains, it Pours: Procyclical Capital Flows and Macroeconomic Policies NBER working paper 10780.

King, Robert, and Charles I. Plosser, 1994, Real Business Cycles and the Test of the Adelmans, Journal of Monetary Economics 33,405-38.

Kose, Ayhan and Raymond Reizman, 2001, "Trade Shocks and Macroeconomic Fluctuations in Africa", Journal of Development Economics, 65, 55-80.

Kose, A. Otrok, C., Whiteman, 2003, International business cycles: world, region, and country specific factors, American Economic Review.93, 1216-1239.

Kose, A., E. Prasad, and M. Terrones, 2005, How Do Trade and Financial Integration Affect the Relationship Between Growth and Volatility, Journal of International Economics, Forthcoming.

Kuznets, Simon, 1958, Long Swings in the Growth of Population and in Related Economic Variables, Proceedings of the American Philosophical Society 102, 25-52.

Kuznets, Simon, 1961, Capital in the American Economy. Its Formation and Financing, Princeton.

Kydland, Finn and Carlos E.J. Zarazaga, 1997, Is the Business Cycle of Argentina "Different"? Economic Review, Federal Reserve Bank of Dallas, Fourth quarter, pp.21-36.

Llona-Rodriguez, Agustin, 2000, Chile During the Gold Standard: A Successful Paper Money Experience, in Aceña, P.M. and Jaime Reis (eds), Monetary Standards in the Periphery. Paper, Silver and Gold, 1854-1913. London.

Maddison, Angus, 1995, Monitoring the World Economy, OECD, Paris.

Maddison, Angus, 2003, The World Economy. Historical Statistics, OECD, Paris.

Marcellino, Massimiliano, 2005, Leading Indicators: What Have We Learned?, Forthcoming in Graham Elliott, Clive Granger and Allan Timmermann (eds.) Handbook of Economic Forecasting. Amsterdam: North-Holland.

Marichal, Carlos, 2002, The Construction of Credibility: Financial Market Reform and the Renegotiation of Mexico's External Debt in the 1880s, in Bortz, Jeffrey and Stephen Haber (eds.) The Mexican Economy, 1870-1930, Stanford, 93-119.

Mendoza, Enrique, 1995, The Terms of Trade, the Real Exchange Rate, and Economic Fluctuations, International Economic Review, 36, 101-37.

Mitchell, Brian, 2003, International Historical Statistics: The Americas, London.

Moore, Geoffrey, 1983, Business Cycle, Inflation, and Forecasting, Cambridge, Mass.

Neumeyer, Pablo and Fabrizio Perri, 2005, Business Cycles in Emerging Economies: The Role of Interest Rates, Journal of Monetary Economics, 52, 345-80.

Pallage, Stephane and Michael Robe, 2003, On the Welfare Cost of Economic Fluctuations in Developing Countries. International Economic Review 44(2), pp. 677- 1.

Romer, Christina, The Prewar Business Cycle Reconsidered: New Estimates of Gross National Product, 1969-1908, Journal of Political Economy, 97, 1, 1-37.

Rostow, W.W. 1980, Why the Poor Gets Richer and the Richer Slows Down: Essays in the Marshallian Long Run, London.

Sheffrin, Steven, 1988, Have Economic Fluctuations Been Dampened? A Look at Evidence Outside the United States, Journal of Monetary Economics, 21,73-83.

Singh, Anoop, Agnes Belaish, Charles Collyns, Paula De Masi, Reva Krieger, Guy Meredith, and Robert Rennhack, 2005, Stabilization and Reform in Latin America, International Monetary Fund Occasional Paper 238.

Solomou, Solomos, 1987, Phases of Economic Growth, 1850-1973. Kondratieff Waves and Kuznets Swings. Cambridge: Cambridge University Press.

Stock, James H., and Mark W. Watson, 1989, New indexes of coincident and leading economic indicators, NBER Macroeconomics Annual, 351-393.

Stock, James H., and Mark W. Watson, 1999, Business Cycle Fluctuations in US Time Series, in John Taylor and Michael Woodford (eds.) Handbook of Macroeconomics, Amsterdam.

Stock, James H., and Mark W. Watson, 2002, Macroeconomic forecasting using diffusion indexes. Journal of Business and Economic Statistics 20:147-162.

Stock, James H., and Mark W. Watson, 2005, Forecasting with Many Predictors, Forthcoming in Graham Elliott, Clive Granger and Allan Timmermann (eds.) Handbook of Economic Forecasting. Amsterdam: North-Holland.

Triner, Gail D. 2000, Banking and Economic Development: Brazil, 1889-1930. New York.

Van Wincoop, Eric, 1999, How Big are the Potential Gains from International Risk Sharing?, Journal of International Economics, 47(1),109-135.

Watson, Mark W., 1994, Business-Cycle Durations and Postwar Stabilization of the U.S. Economy, American Economic Review, 84 (1), 24-46.

4.6 Appendix A: Data Construction and Sources

4.6.1 Argentina

Agricultural Output

1900-1960: Banco Central de Argentina (1976)

1961-2004: World Bank, World Development Database.

Industrial Output

1875-1960: della Paolera and Taylor (2003)

1961-2004: World Bank, World Development Database.

Transport Output

1870-1960: Geometric weighted average of passengers and tons of freight per kilometers times total railway road extension; then spliced in 1913 with the index provided in Carlos F. Diaz-Alejandro, 1970, Essays on the Economic History of the Argentine Republic. Yale.

Cement Consumption

1870-1913: Total cement imports in tons from the United Kingdom and the United States, which together accounted for no less than between 60 to 70 percent of Argentina's total cement imports. The sources are the United Kingdom, Board of Trade, Annual Statements of the Trade of the United Kingdom with Foreign Countries and British Possessions. London: HMSO. Commerce and Navigation, several issues; and the United States, Foreign Commerce, Navigation and Tonnage of the United States, Washington, DC: Department of Labor, several issues. Because a local cement industry was non-existing before World War I, all domestic consumption of cement was then met by imports. So this newly constructed series for the period should be expected to a good proxy for domestic construction activity.

1913-2000: Oxford Latin American Economic History database, available at http://oxlad.qeh.ox.ac.uk/.

2001-2004: Instituto Nacional de Estadisticas y Census de la Republica Argentina, available at: www.indec.mecon.ar/.

Fixed Investment

1870-1884: Capital goods imports from the United Kingdom and the United States (converted into equivalent pounds sterling) and deflated by the UK capital good deflator taken from Charles H. Feinstein, 1972, Statistical Tables of National Income, Expenditure and Output of the United Kingdom, 1855-1965, Cambridge. Since the domestic capital goods industry was virtually non-existent in Argentina before World War I (being in fact relatively negligible before WWII – see Diaz-Alejandro, 1970), and because the UK and the US were the two most important suppliers of capital goods to Argentina, such imported capital goods series should be expected proxy very well aggregate fixed capital formation in the country in those early decades.

1885-1960: della Paolera and Taylor (2003).

1961-2004: International Financial Statistics (IFS), International Monetary Fund.

Central Government Expenditures and Revenues

1870-1960: Brian M. Mitchell, 1998, International Historical Statistics: The Americas, London.

1961-2004: Luis A.V. Catão and Marco E. Terrones, 2005, "Fiscal Deficits and Inflation," Journal of Monetary Economics, 52, 529-554. Both series are expressed in real terms by deflating them by the consumer price index (CPI).

Narrow (M0) and Broad Money (M2)

1870-1960: Mitchell, op cit.

1961-2004: IFS. Both series expressed in real terms by deflating them by the CPI.

Consumer Price Index (CPI)

1870-1913: Catão and Solomou (2005).

1914-1960: della Paolera and Taylor (2003)

1961-2004: IFS.

Average Interest Rate on Domestic Public Bonds

1870-1913: 'Monetary and Banking Experiments in Argentina: 1861-1930', Paper presented at the conference, 'Economic Growth in the Long Run: Argentina, Brazil and Mexico, 1870-1950' at the Institute of Latin American Studies, University of London, March, 1992.

1914-1993: della Paolera and Javier Ortiz, 1995, Dinero, Intermediacion Financiera y nivels de actividad en 110 anos de historia economica Argentina. Documentos de Trabajo 36 (December), Universida Torçuato di Tella.

1993-2004: IFS (line 60p). Real interest rate series obtained by deflating annual nominal yields by current period CPI inflation.

Export and Import volumes and Net Barter Terms of Trade

1870-1913: Catão and Solomou (2005).

1914-1960: Oxford Latin American Economic History database, available at http://oxlad.qeh.ox.ac.uk/.

1961-2004: IFS.

Real Effective Exchange Rate

1870-1913: Catão and Solomou (2005).

1914-2004: CPI-based geometric weighted averages of Argentina's real bilateral exchange rates with its eight largest trading partners (covering between 67 and 80 percent of visible trade).[20] Fisher ideal indices were derived for the sub-periods 1914-1946 and 1946-2004 (based on 1913 and 1938, and 1960 and 2000 weights respectively), and then spliced at 1946. Nominal exchange rates for the entire post-war period are market rates underlying Carmen M. and Kenneth S. Rogoff, 2004, "The Modern History of Exchange Rate Arrangements: A Reinterpretation," Quarterly Journal of Economics, CXIX, No.1, pp.1-48.

Net Foreign Capital Inflows

1870-1960: Obtained by splicing the series on UK capital flows to Argentina provided in Stone, Irving, 1999, The Global Export of Capital from Great Britain, 1865-1914: A Statistical Survey, New York, with a post-1884 series on net capital inflows constructed as changes in end-year net international reserves

[20]The choice of GDP deflator rather than a CPI-based index was determined by the deficiencies of the existing CPI series during the period 1870-1913, compared to an existing series based on production weights (therefore mimicking a GDP deflator) which covers a much extensive range of products and constructed based on weights from national production censuses.

expressed in US$ million (obtained from Gerardo della Paolera, 1988, "How the Argentine Economy Performed During the International Gold Standard: A Re-examination", PhD thesis, University of Chicago for 1870-1913 then with the Cavallo-Mundlak series, as kindly supplied by Alan Taylor) minus the current account balance (also expressed in US$ millions) provided in della Paolera and Taylor (2003). The splicing of the two series is warranted by the fact that the UK was by far the most important souce of foreign capital flows to Argentina before World War I (and particularly prior to 1890), and evidence that the two series co-move tightly together in the 1884-1913 period, with a correlation coefficient of 0.81.

1961-2004: Also obtained as the difference between changes in international reserves and the current account balance, both as reported by the IFS. The resulting nominal series in US dollars was then deflated by the US Wholesale price index (WPI) obtained from Global Financial Database for the period 1870-1947 and the IFS for 1948-2004.

Wages

1870-1913: Jeffrey G. Williamson, "The Evolution of Global Labor Markets Since 1830", Explorations in Economic History, 32 (2), 1995, pp. 141-96.

1914-1980: della Paolera and Taylor (2003).

1981-2004: IMF's WEO database. This series was then deflated by CPI to obtain the real wage index.

Foreign 3-month bill rate

1870-1920: Annual average yields of 3-month bills on the London market provided in Sidney Holmer and Richard Sillas, 1996, A History of Interest Rates, Rutgers.

1921-2004: Annual average yields of the US 3-month Treasury Bill provided in the same source. The choice of 1920 as the splicing point was due to the unavailability of the US instrument prior to 1920.

Both series were deflated by the respective countries' CPI inflation, obtained from Catao and Solomou (2005) for 1870-1913, Mitchell, op cit (1914-1960) and the IFS (1961-2004).

Foreign Output

Sum of France's, Germany's, UK's and US's GDP, all expressed in 1990 PPP constant dollars from Maddison (2003).

Population at mid-year

1870-1960: Maddison (2003).

1961-2004: IFS (2003).

4.6.2 Brazil

Agricultural and Manufacturing Output

1900-1946: Haddad (1978).

1947-2004: Instituto Brasileiro de Geografia e Estatistica, available at http://www.IBGE.gov.br.

Transport Output

1870-1907: Average of freight and passenger transported in railways, using 1908 weights provided in Haddad (1978).

1908-1946: Haddad (1978).

1947-2004: Instituto Brasileiro de Geografia e Estatistica, available at http://www.IBGE.gov.br.

Communications Output

1870-1907: Average of mail and telegraph trafic in the national postal system, weighted according to current 1889 values provided in Instituto Brasileiro de Geografia e Estatistica, 1987. Estatisticas Históricas do Brasil, Rio de Janeiro, IBGE.

1908-1946: Haddad (1978).

1947-2004: Instituto Brasileiro de Geografia e Estatistica, available at http://www.IBGE.gov.br.

Cement Consumption

1870-1901: Cement imports from the France, Germany, the UK, and the US, obtained from these countries' own trade statement data (see above). Since these four countries accounted for between 75 and 85 percent of total Brazilian imports (see IBGE, op.cit., pp. 545-49) and all cement consumed in Brazil at the time was imported, this newly constructed series is very representative of aggregate cement consumption and hence a good proxy for domestic construction activity.

1902-1945: Anibal V. Villela and Wilson Suzigan, 1975, Politica do Governo e Crescimento da Economia Brasileira 1889-1945, p.423.

1945-2004: Instituto Brasileiro de Geografia e Estatistica, available at http://www.IBGE.gov.br; and IPEA database.

Machinery Investment

1870-1913: Wilson Suzigan, Industria Brasileira. Origens e Desenvolvimento, São Paulo, 1986.

1913-2004: Instituto Brasileiro de Geografia e Estatistica, available at http://www.IBGE.gov.br.

Central Government Expenditures and Revenues

1870-1960: IBGE, op. cit.

1961-2004: Luis A.V. Catão and Marco E. Terrones, 2005, "Fiscal Deficits and Inflation," Journal of Monetary Economics, 52, 529-554. Both series are expressed in real terms by deflating them by the GDP deflator.

Narrow (M1) and Broad Money (M2)

1870-1960: IBGE, op. cit.

1961-2004: IFS. Both series expressed in real terms by deflating them by the GDP deflator.

GDP deflator

1870-1913: Catão and Solomou (2005).

1914-1960: IBGE, op cit.

1961-2004: IFS.

Domestic Interest Rate

1870-1961: yields on government perpetuities ("apolices"). Pre-1930 data kindly provided by Bill Summerhill and Gail Trinner based on their research with Brazilian historical archieves. Post-1930 data from Claudio Contador, 1975, O Mercado de Ativos Financeiros no Brasil. Perspectiva historica e comportamento recente. Rio de Janeiro.

1965-1980: Equivalent nominal yield on inflation indexed public bonds (ORTNs), from Goldsmith, op. cit. The gap between the apolice series and the ORTN series was bridged by linear interpolation.

1981-2004: Money market interest rate from IFS. Real interest rate series obtained by deflating annual nominal yields by current annual percentage changes in the GDP deflator.

External Interest Rate Spread

1870-1913: Mauro, Paolo, N. Sussman and Y. Yafeh, 2002, "Emerging Market Spreads: Then versus Now," Quarterly Journal of Economics, CXVII.

1914-1956: Global Financial Data.

1968-1986: Eliana Cardoso and Albert Fishlow, "The Macroeconomics of Brazilian External Debt," in J. Sachs (ed.), Developing Country Debt and Economic Performance. Chicago, 269-391

1987-1993: Estimated as the one-year libor interest rate plus a 400 basis points spread minus the US 10-year bond interest rate.

1993-2004: IMF's global data source database. Real interest rate series obtained by deflating annual nominal yields by current period CPI inflation.

Export and Import volumes and Net Barter Terms of Trade

1870-1913: Catão and Solomou (2005).

1914-1960: Instituto Brasileiro de Geografia e Estatistica, available at http://www.IBGE.gov.br

1961-2004: IFS.

Real Effective Exchange Rate

1870-1913: Catão and Solomou (2005).

1914-2004: GDP deflator-based geometric weighted averages of Brazil's real bilateral exchange rates with its eight largest trading partners (covering between 64 and 75 percent of visible trade). Fisher ideal indices were derived for the sub-periods 1914-1946 and 1946-2004 (based on 1913 and 1938, and 1960 and 2000 weights respectively), and then spliced at 1946. Nominal exchange rates for the whole post-WWII period are market rates underlying Carmen M. and Kenneth S. Rogoff, 2004, "The Modern History of Exchange Rate Arrangements: A Reinterpretation," Quarterly Journal of Economics, CXIX, No.1, pp.1-48.

Wages

1870-1940: Jeffrey G. Williamson, "The Evolution of Global Labor Markets Since 1830", Explorations in Economic History, 32 (2), 1995, pp. 141-96.

1940-1955: IBGE, op cit..

1955-1976: Raymond Goldsmith, 1986, O Desenvolvimento Financeiro do Brasil, Sao Paulo.

1977-2004: IBGE, op cit..

Foreign 3-month bill rate and Foreign Output

The same as for Argentina.

Population at mid-year

1870-1960: Maddison (2003).

1961-2004: IFS (2003).

4.6.3 Chile

Agriculture, Manufacturing and Mining Output

1870-1960: Juan Braun, Matías Braun, Ignacio Briones, José Diaz, Rolf Luders and Gert Wagner, 2000, "Economía Chilena 1810–1995: Estadísticas Históricas", Documento de Trabajo No. 187, Catholic University of Chile, Instituto de Economía.

1961-2004: World Development database, World Bank, and Central Bank of Chile.

Machinery Investment

1870-1900: Capital goods imports from the United Kingdom and the United States (converted into equivalent pounds sterling) and deflated by the UK capital good deflator taken from Charles H. Feinstein, 1972, Statistical Tables of National Income, Expenditure and Output of the United Kingdom, 1855-1965, Cambridge.

1901-1990: Andre Hoffman, 2000, The Economic Development of Latin America in the Twentieth Century, Cheltenham (pp. 190-91, Table D2).

1991-2004: Central Bank of Chile, ibid.

Central Government Expenditures and Revenues

1870-1960: Brian M. Mitchell, 1998, International Historical Statistics: The Americas, London.

1961-2004: Luis A.V. Catão and Marco E. Terrones, 2005, "Fiscal Deficits and Inflation," Journal of Monetary Economics, 52, 529-554. Both series are expressed in real terms by deflating them by the consumer price index (CPI).

Narrow Money (M0) and Broad Money (M2)

1870-1878: Mo calculated as paper money issued minus banks' cash-in-vault, both taken from Llona Rodriguez, Agustin, Chilean Monetary Policy 1870-1925, PhD thesis, Boston University; M2 from Braun et al, op.cit.

1879-1960: Mo from Mitchell, op cit.; M2 from Braun et al. op.cit.

1961-2004: IFS. Both series expressed in real terms by deflating them by the CPI.

Mortgage Credit

1870-1995: Braun et al. (1995).

1995-2004: Central Bank of Chile.

Consumer Price Index (CPI)

1870-1913: Catão and Solomou (2005).

1914-1960: Mitchell, op cit.

1961-2004: IFS.

Domestic Interest Rate

1870-1993: Bank lending rate from Braun et al, op cit.

1993-2004: IFS (line 60p). Real interest rate series obtained by deflating annual nominal yields by current period CPI inflation.

Export and Import volumes and Net Barter Terms of Trade

1870-1913: Catão and Solomou (2005).

1914-1960: Braun et al, op.cit.

1961-2004: IFS.

Real Effective Exchange Rate

1870-1913: Catão and Solomou (2005).

1914-2004: CPI-based geometric weighted averages of real bilateral exchange rates with up to Chile's twenty largest trading partners (covering between 79 and 93 percent of visible trade). Fisher ideal indices were derived for the sub-periods 1914-1946 and 1946-2004 (based on 1913 and 1938, and 1960 and 2000 weights respectively), and then spliced at 1946. Nominal exchange rates for the whole post-WWII period are market rates underlying Carmen M. and Kenneth S. Rogoff, 2004, "The Modern History of Exchange Rate Arrangements: A Reinterpretation," Quarterly Journal of Economics, CXIX, No.1, pp.1-48.

Wages

1870-1995: Real wage index from Braun et al, op.cit.

1995-2004: Average nominal wage index from IMF's WEO database, deflated by CPI.

Population at mid-year

1870-1960: Maddison (2003).

1961-2004: IFS (2003).

4.6.4 Mexico

Agricultural Output

1878-1910: Colégio de México, 1960, Estadísticas Economicas del Porfiriato, p.61. Refers to export crop sub-index. Converted from a fiscal to calendar year basis by the averaging of two successive years.

1911-1921: Index constructed as a weighted average of the output of ten main crops (beans, corn (maiz), cotton, coffee, garbanzo, rice, sisal, sugar, and tomatoes) weighted by their 1900 (normalized) share in total value of agricultural production. The information on individual crop output was taken from INEGI, 1992, Estadísticas Historicas de Mexico, Mexico.

1921-1960: Leopoldo Solis, 1975, La Realidade Económica Mexicana. Retrovision y Perspectivas. Mexico.

1961-2004: World Development database, World Bank, and INEGI, available at http://www.inegi.gob.mx/est/default.asp?c=1601.

Manufacturing Output

1878-1910: Colégio de México, op. cit, p.105. Prior to 1892, the series reflects solely changes in the index of domestic textile production taken from Haber, Stephen, Armando Razo and Noel Maurer, 2003, The politics of property rights: Political instability, credible commitments, and economic growth in Mexico, 1876-1929. Cambridge. Figures for 1879-1882, 1884-87, and 1890 derived by linear interpolation due to the gaps in the original source.

1910-1921: INDEC, op.cit.

1921-1960: Solis, 1975, op.cit.

1961-2004: World Development database, World Bank, and INEGI, available at http://www.inegi.gob.mx/est/default.asp?c=1601.

Mining Ouput

1878-1910: Colégio de México, op. cit, p.135.

1911-1921: Weighted average of the output of ten main domestically produced metals (silver, gold, iron, graffite, lead, mercury, copper, zinc, antimonio, and lead) as well as oil, weighted according their 1900 value share in total mining output provided in the same source (pp.136-43).

1921-1960: Solis, 1975, op.cit.

1961-2004: World Development database, World Bank, and INEGI, available at http://www.inegi.gob.mx/est/default.asp?c=1601.

Transportation and Communications Output

1870-1921: Weighted average of railway freight and passanger traffic (taken from John Coatsworth, Growth Against Development – The Economic Impact of Railways in Porfirian Mexico, Illinois) and postal service traffic, taken from Mitchell, op.cit.

1921-1960: Solis, 1975, op.cit.

1961-2004: World Development database, World Bank, and INEGI, available at http://www.inegi.gob.mx/est/default.asp?c=1601.

Cement Consumption

1870-1931: Cement imports from the UK and the US (by far the two main foreign suppliers), obtained from these countries' own trade statement data (see above). From 1906 onwards, when the first plants of domestic cement production began operations, we add their output (taken from the Oxford Latin American Economic History database, see above) to imports.

1932-2000: Oxford Latin American Economic History database, available at http://oxlad.qeh.ox.ac.uk/.

2001-2004: INEGI, op.cit.

Machinery Investment

1870-1925: Luis Catão, 2005, "Exchange Rate and Machinery Investment in Early Development: The Cases of Brazil and Mexico." International Monetary Fund. Mimeo,.

1925-1940: Enrique Cardenas, 1987, Mexico's Industrialization during the Great Depression: Public Policy and Private Response, PhD thesis, Yale University. The index is based on the volume of imported capital goods.

1941-1993: Hoffman, op.cit.

1994-2004: INEGI, op.cit.

Central Government Expenditures and Revenues

1870-1960: IBGE, op. cit.

1961-2004: Luis A.V. Catão and Marco E. Terrones, 2005, "Fiscal Deficits and Inflation," Journal of Monetary Economics, 52, 529-554. Both series are expressed in real terms by deflating them by the CPI deflator.

Narrow (M0) and Broad Money (M2)

1870-1925: Catão, op. cit.

1926-1960: INEGI, op.cit.

1961-2004: IFS. Resulting real series was obtained by CPI deflation.

CPI

1870-1913: Catão and Solomou (2005).

1914-1917: Interpolated assuming relative PPP, given that no domestic data seems available for the hyperinflation period. Assuming PPP is probably not very inaccurate since those years were characterizing by soaring inflation and a hyperinflation (see main text) which typically tends to align domestic price movements with the exchange rate.

1918-1940: Williamson, op.cit.

1941-1960: Mitchell, op.cit.

1961-2004: IFS.

Export and Import volumes and Net Barter Terms of Trade:

1870-1925: Catão (2005), op.cit.

1926-1940: Cardenas, op.cit.

1941-1960: INEGI, op.cit.

1961-2004: IFS.

Real Effective Exchange Rate

1870-1913: Catão and Solomou (2005).

1914-2004: CPI-based geometric weighted averages of Mexico's real bilateral exchange rates with its six largest trading partners (covering over 90 percent of visible trade). Fisher ideal indices were derived for the sub-periods 1914-1946 and 1946-2004 (based on 1913 and 1938, and 1960 and 2000 weights respectively), and then spliced at 1946. Nominal exchange rates for the whole post-WWII period are market rates underlying Carmen M. and Kenneth S. Rogoff, 2004, "The Modern History of Exchange Rate Arrangements: A Reinterpretation," Quarterly Journal of Economics, CXIX, No.1, pp.1-48.

Wages

1870-1940: Jeffrey G. Williamson, "The Evolution of Global Labor Markets Since 1830", Explorations in Economic History, 32 (2), 1995, pp. 141-96.

1940-1974: Mitchell, op cit..

1974-2004: WEO and IFS databases.

Foreign 3-month bill rate

1870-1920: Annual average yields of 3-month bills on the London market provided in Sidney Holmer and Richard Sillas, 1996, A History of Interest Rates, Rutgers.

1921-2004: Annual average yields of the US 3-month Treasury Bill provided in the same source. The choice of 1920 as the splicing point was due to the unavailability of the US instrument prior to 1920. Both series were deflated by the respective countries' CPI inflation, obtained from Catao and Solomou (2005) for 1870-1913, Mitchell, op cit (1914-1960) and the IFS (1961-2004).

Foreign Output

US real GDP in 1990 PPP constant dollars from Maddison (2003).

Population at mid-year

1870-1960: Maddison (2003).

1961-2004: IFS (2003).

4.6.5 United States

Agricultural Output

1870-1949: NBER historical database (output of main crops)

1950-2004: Chained quantity index of value added in agriculture, Bureau of Economic Analysis .

Manufacturing Output

1870-1949: Balke and Gordon (1989), kindly communicated by the authors.

1950-2004: Chained quantity index of value added in manufacturing, Bureau of Economic Analysis .

Transportation and Communications Output

1870-1949: Balke and Gordon (1989), kindly communicated by the authors.

1950-2004: Chained quantity index of value added in tranportation and information industries, Bureau of Economic Analysis .

Construction Activity

1870-1949: Kuznets' construction series spliced with the Gottlieb/BLS counterpart, as reported in Balke and Gordon (1989), kindly communicated by the authors.

1950-2004:Chained quantity index of value added in construction, Bureau of Economic Analysis

Machinery Investment

1870-1949: Kuznets' durable producers' output series (with 5-year moving average unravelled), kindly provided by Christina Romer..

1950-2004: Bureau of Economic Analysis.

Central Government Expenditures and Revenues

1870-1958: Mitchell, op cit..

1959-2004: IFS.

Narrow (M0) and Broad Money (M2)

1870-1970: Bureau of Economic Analysis, Historical Statistics of the United States.

1970-2004: Haver Analytics.

CPI

1870-1913: Catão and Solomou (2005).

1914-1947: Balke and Gordon (1989)

1948-2004: IFS.

Export and Import volumes and Net Barter Terms of Trade:

1870-1913: Catão and Solomou (2005), op.cit.

1914-1959: Bureau of Economic Analysis, Historical Statistics of the United States.

1960-2004: IFS.

Current Account Balance:

1870-1959: Bureau of Economic Analysis, Historical Statistics of the United States. Calculated by adding the balance of goods and services (series U15) plus net private and government transfers (series U16 and U17).

1960-2004: IFS.

Real Effective Exchange Rate

1870-1913: Catão and Solomou (2005).

1914-2004: CPI-based geometric weighted averages of the United States's real bilateral exchange rates with 17 largest trading partners (covering over no less than two thirds of its visible trade before World War II and no less than 80 percent onwards). Fisher ideal indices were derived for the sub-periods 1914-1946 and 1946-2004 (based on 1913 and 1938, and 1970 and 2000 weights respectively), and then spliced at 1946. Nominal exchange rates for the whole post-WWII period are market rates underlying Carmen M. and Kenneth S. Rogoff, 2004, "The Modern History of Exchange Rate Arrangements: A Reinterpretation," Quarterly Journal of Economics, CXIX, No.1, pp.1-48.

Wages

1870-1944: Jeffrey G. Williamson, "The Evolution of Global Labor Markets Since 1830", Explorations in Economic History, 32 (2), 1995, pp. 141-96.

1945-2004: Hourly compensation in manufacturing from IFS deflated by CPI from the same source.

Foreign 10-year bill rate

1870-1953: Sidney Holmer and Richard Sillas, 1996, A History of Interest Rates, Rutgers.

1954-2004: IFS.

Foreign Output

Real GDP of Australia, Canada, UK, France, Germany, Italy and Japan in 1990 PPP constant dollars from Maddison (2003) extended with IFS data through 2004.

Population at mid-year

1870-1960: Maddison (2003).

1961-2004: IFS (2003).

Table 4.1: Factor Loadings

The table reports the eigenvectors associated with the first two common factors constructed from the data sample covering the period 1870-2004.

	Argentina		Brazil		Chile		Mexico	
	F1	F2	F1	F2	F1	F2	F1	F2
Agriculture	-	-	-	-	0.03	-0.21	0.25	-0.22
Communication	-	-	0.29	-0.11	-	-	-	-
Manufacturing	-	-	-	-	0.22	0.29	0.19	0.21
Mining	-	-	-	-	0.35	-0.16	0.26	0.05
Cement	0.30	-0.06	0.34	-0.07	-	-	0.21	0.24
Transportation	0.35	-0.05	0.21	-0.26	-	-	0.26	-0.05
Fixed Investment	0.37	-0.09	0.42	0.02	0.34	-0.03	0.30	0.30
Govt. Expenditures	0.30	0.31	0.21	0.36	0.30	0.08	0.24	-0.26
Govt. Revenues	0.31	-0.04	0.31	0.22	0.34	-0.03	0.27	-0.16
Narrow Money	0.19	0.39	0.11	0.18	0.11	0.45	0.34	-0.13
Broad Money	0.30	0.16	0.05	0.18	0.18	0.42	0.33	-0.14
Inflation	-0.08	0.30	-0.10	-0.04	-0.07	0.03	-0.18	-0.03
Domestic Interest Rate	-0.05	-0.37	-0.07	-0.02	-0.00	-0.33	-	-
Mortgage Credit	-	-	-	-	0.05	-0.20	-	-
Terms of Trade	0.17	0.09	0.27	-0.31	0.19	0.06	0.09	0.28
Real Exchange Rate	0.00	-0.43	0.20	-0.19	0.10	-0.43	0.14	0.43
Export Volume	0.04	-0.11	-0.05	0.17	0.31	-0.25	0.21	-0.12
Import Volume	0.36	-0.26	0.43	-0.06	0.39	0.01	0.28	0.38
Trade Balance	-0.25	0.24	-0.30	-0.11	-0.06	-0.16	-0.08	-0.22
External Spread	-	-	0.03	0.27	-	-	-	-
Foreign Capital Inflows	0.23	-0.17	-	-	-	-	-	-
Foreign 3m-Tbill	0.08	-0.11	0.08	0.10	-0.06	0.13	0.17	-0.25
Foreign Output	-0.01	0.11	0.02	-0.43	0.26	-0.00	0.09	0.16
Real wage	0.20	0.25	0.02	0.40	0.29	-0.06	0.25	-0.24
Population	0.09	0.20	-0.10	-0.27	0.01	0.12	0.00	-0.10

Table 4.2: In-Sample Fit

The table reports the adjusted R-squared for the backcasting equation estimated over the post-war sample (1950-2004). Panel A reports results when the factors are extracted from a panel spanning the period 1870-2004 (1878-2004 for Mexico), while in panel B the factors are extracted from a larger cross-section of variables available during the sample 1900-2004. All-regressors reports results from the backcasting equation that includes all available variables. The remaining backcasting equations estimate the factors using either the Stock and Watson (2002) approach with r static factors (SW(r)) or the Forni et al (2000) dynamic factor approach with q, r dynamic and static factors, respectively (FHLR(q,r)).

Panel A: 1870-2004

	Argentina	Brazil	Chile	Mexico
All regressors	0.89	0.77	0.91	0.92
SW(1)	0.76	0.51	0.72	0.83
SW(2)	0.81	0.52	0.74	0.83
SW(3)	0.81	0.69	0.75	0.86
SW(4)	0.82	0.71	0.78	0.87
FHLR(1,1)	0.79	0.47	0.71	0.75
FHLR(1,2)	0.82	0.67	0.71	0.76
FHLR(1,3)	0.81	0.68	0.71	0.87
FHLR(1,4)	0.81	0.67	0.80	0.86
FHLR(2,1)	0.80	0.46	0.67	0.77
FHLR(2,2)	0.81	0.54	0.74	0.76
FHLR(2,3)	0.80	0.56	0.74	0.84
FHLR(2,4)	0.80	0.68	0.80	0.85

Panel B: 1900-2004

	Argentina	Brazil	Chile	Mexico
All regressors	0.94	0.90	0.91	0.92
SW(1)	0.84	0.65	0.72	0.83
SW(2)	0.88	0.65	0.74	0.83
SW(3)	0.88	0.83	0.75	0.86
SW(4)	0.88	0.84	0.78	0.87
FHLR(1,1)	0.85	0.61	0.71	0.75
FHLR(1,2)	0.87	0.79	0.71	0.76
FHLR(1,3)	0.88	0.80	0.71	0.87
FHLR(1,4)	0.88	0.80	0.80	0.86
FHLR(2,1)	0.86	0.60	0.67	0.77
FHLR(2,2)	0.88	0.66	0.74	0.76
FHLR(2,3)	0.88	0.70	0.74	0.84
FHLR(2,4)	0.88	0.81	0.80	0.85

Table 4.3: In-sample and Out-of-sample Fit for the USA

The table reports the adjusted R-squared for the backcasting equation estimated over the post-war sample (1950-2004) and the pre-war period (1870-1949) with US data. The factors are extracted from a panel spanning the period 1870-2004. For the post-war sample we report the results using the Balke and Gordon series while for the pre-war period we report results obtained by using the Balke-Gordon, Romer, and Kuznets estimates.

	1950-2004	1870-1949		
	Balke-Gordon	Balke-Gordon	Romer	Kuznets
All regressors	0.962	0.893	0.894	0.811
SW(1)	0.762	0.725	0.681	0.675
SW(2)	0.758	0.725	0.677	0.730
SW(3)	0.794	0.863	0.832	0.807
SW(4)	0.800	0.874	0.866	0.805
FHLR(1,1)	0.761	0.763	0.716	0.601
FHLR(1,2)	0.761	0.764	0.723	0.723
FHLR(1,3)	0.803	0.847	0.823	0.740
FHLR(1,4)	0.802	0.846	0.822	0.737
FHLR(2,1)	0.816	0.802	0.766	0.727
FHLR(2,2)	0.813	0.809	0.767	0.766
FHLR(2,3)	0.809	0.834	0.804	0.764
FHLR(2,4)	0.819	0.887	0.868	0.790

Table 4.4: Dating the Cycle

The table reports peak and trough dates selected by the Bry-Boschan algorithm. Results in Panel A impose a minimum of two years between peaks, while results in Panel B impose a minimum of six years between peaks.

Panel A

Argentina		Brazil		Chile		Mexico	
Peaks	Troughs	Peaks	Troughs	Peaks	Troughs	Peaks	Troughs
1873	1880	1872	1876	1875	1877	1883	1886
1884	1887	1881	1886	1882	1885	1891	1894
1889	1891	1891	1893	1890	1894	1898	1902
1893	1895	1896	1901	1896	1897	1907	1909
1899	1902	1907	1908	1898	1902	1912	1916
1913	1918	1912	1915	1907	1909	1921	1927
1923	1926	1927	1931	1912	1915	1928	1932
1930	1933	1938	1942	1918	1919	1936	1938
1938	1945	1946	1948	1926	1927	1941	1942
1948	1953	1951	1956	1929	1932	1945	1947
1958	1959	1962	1967	1937	1942	1951	1953
1961	1963	1974	1976	1943	1947	1956	1959
1965	1968	1979	1983	1952	1954	1960	1962
1970	1971	1986	1988	1957	1958	1965	1971
1974	1978	1990	1992	1962	1965	1974	1977
1980	1985	1997	1999	1966	1970	1981	1988
1987	1990	2002	-	1972	1975	1994	1995
1994	1995	-	-	1980	1983	2000	2002
1997	2002	-	-	1989	1990	-	-
-	-	-	-	1997	2002	-	-

Panel B

Argentina		Brazil		Chile		Mexico	
Peaks	Troughs	Peaks	Troughs	Peaks	Troughs	Peaks	Troughs
1873	1880	1881	1876	1890	1885	1883	1886
1889	1891	1891	1886	1912	1909	1898	1916
1899	1902	1913	1901	1929	1919	1921	1927
1913	1918	1927	1915	1943	1932	1936	1938
1930	1933	1938	1931	1952	1947	1951	1959
1948	1953	1951	1942	1980	1958	1965	1971
1961	1971	1962	1956	1997	1983	1981	1995
1980	1985	1979	1967	-	-	-	-
1997	-	1997	1992	-	-	-	-

Table 4.5: Spectral Density Function Estimates of Cyclical Durations

The table reports business cycle durations (in years) for different sample periods. The estimates refer to the peak value of a Bartlett lag window estimate using a bandwidth set at twice the number of sample observations.

	1870-1929		1930-1970		1971-2004	
	Estimate	T	Estimate	T	Estimate	T
Argentina	4.64	16	3.36	12	2.11	12
	(1.96)		(1.48)		(1.02)	
Brazil	4.57	16	3.90	12	2.35	12
	(1.92)		(1.72)		(1.13)	
Chile	2.18	16	2.05	12	3.79	8
	(0.92)		(0.90)		(1.90)	
Mexico	3.75	14	2.35	8	3.14	12
	(1.59)		(1.04)		(1.52)	

Table 4.6: Persistence and Volatility of the Cycle

The table reports autocorrelation and standard deviation estimates of the cycle obtained from a static two-factor model. Panel A shows results for the backcasted data while Panel B reports results for actual business cycle measures.

Panel A: Backcasted Data

	1870-1929				1930-1970			
	Arg	**Bra**	**Chi**	**Mex**	**Arg**	**Bra**	**Chi**	**Mex**
AR(1)	0.721	0.706	0.549	0.739	0.593	0.763	0.538	0.573
Std	0.065	0.048	0.063	0.066	0.046	0.032	0.078	0.029
	1971-1987				1988-2004			
	Arg	**Bra**	**Chi**	**Mex**	**Arg**	**Bra**	**Chi**	**Mex**
AR(1)	0.089	0.684	0.614	0.606	0.579	0.402	0.562	0.286
Std	0.029	0.037	0.071	0.039	0.065	0.020	0.020	0.024

Panel B: Actual Data

	1971-1987				1988-2004			
	Arg	**Bra**	**Chi**	**Mex**	**Arg**	**Bra**	**Chi**	**Mex**
AR(1)	0.168	0.624	0.590	0.644	0.590	0.622	0.687	0.383
Std	0.034	0.043	0.081	0.039	0.068	0.021	0.033	0.030

Table 4.7: International Comparisons

The table reports autocorrelation and standard deviation estimates of the business cycle. Panel A shows median regional values for Latin America, LA-4, (Argentina, Brazil, Chile and Mexico), European Countries, EU-4, (France, Germany, Italy and the UK), New World countries, NW-3, (USA, Canada and Australia) and Japan, JP. The estimates of the cycle for Latin America are obtained from a dynamic factor model. Panel B compares median regional values for Latin America, LA-4, (Argentina, Brazil, Chile and Mexico), other developing countries, Other LDCs, (India, Indonesia, Korea, Malaysia, Taiwan, South Africa, and Turkey), Africa (41 countries), Asia (11 countries), and the Middle East (18 countries) based on actual data.

Panel A: Advanced Countries Data

	1870-1929				1930-1970			
	LA-4	EU-4	NW-3	JP	LA-4	EU-4	NW-3	JP
AR(1)	0.719	0.507	0.517	0.031	0.583	0.698	0.731	0.599
Std	0.065	0.051	0.050	0.035	0.039	0.101	0.066	0.128
	1971-1987				1988-2004			
	LA-4	EU-4	NW-3	JP	LA-4	EU-4	NW-3	JP
AR(1)	0.610	0.526	0.477	0.683	0.482	0.619	0.501	0.417
Std	0.038	0.017	0.021	0.027	0.022	0.015	0.022	0.018

Panel B: Developing Countries Data

	1930-1950					1950-1970				
	LA-4	Other LDCs	Africa	Asia	Middle East	LA-4	Other LDCs	Africa	Asia	Middle East
AR(1)	0.560	0.617	-	-	-	0.273	0.051	0.502	0.416	0.402
Std	0.040	0.078	-	-	-	0.031	0.030	0.038	0.028	0.043
	1971-1987					1988-2004				
	LA-4	Other LDCs	Africa	Asia	Middle East	LA-4	Other LDCs	Africa	Asia	Middle East
AR(1)	0.607	0.551	0.563	0.486	0.447	0.606	0.531	0.481	0.547	0.572
Std	0.041	0.038	0.051	0.033	0.031	0.031	0.028	0.035	0.028	0.057

Table 4.8: Cyclical Volatility Estimates of Selected Variables

The table reports standard deviations (in percent) for selected variables and different sample periods.

Panel A: 1870-1929

	Argentina	Brazil	Chile	Mexico
Foreign Interest Rate	5.34	5.34	5.34	5.34
Foreign Output	3.46	3.46	3.46	3.46
Domestic Output	6.49	4.84	6.25	7.29
Terms of Trade	14.99	21.79	33.05	29.00
Export Volume	13.20	10.21	17.61	10.34
Fixed Investment	107.80	37.16	76.05	79.00
Government Spending	66.23	47.85	39.33	33.71
Expenditure/Revenue	28.41	20.82	28.83	10.70
Narrow Money	14.94	15.32	16.64	35.07
Broad Money	14.82	14.62	11.39	35.83
Real Wage	12.33	9.14	8.69	32.73

Panel B: 1930-1970

	Argentina	Brazil	Chile	Mexico
Foreign Interest Rate	4.68	4.68	4.68	4.68
Foreign Output	7.08	7.08	7.08	7.08
Domestic Output	4.56	3.25	7.69	2.90
Terms of Trade	12.20	20.27	26.66	29.37
Export Volume	13.59	9.95	16.89	9.52
Fixed Investment	67.96	29.56	64.60	102.36
Government Spending	14.09	13.27	13.69	16.73
Expenditure/Revenue	23.36	16.21	13.83	19.64
Narrow Money	10.20	11.01	19.90	15.43
Broad Money	12.33	10.84	13.46	13.62
Real Wage	6.99	6.54	13.07	18.22

Panel C: 1971-2004

	Argentina	Brazil	Chile	Mexico
Foreign Interest Rate	2.20	2.20	2.20	2.20
Foreign Output	2.97	2.97	2.97	2.97
Domestic Output	5.01	2.94	5.35	3.30
Terms of Trade	27.14	12.24	16.45	22.20
Export Volume	12.94	8.11	6.87	9.61
Fixed Investment	16.39	15.82	16.10	14.93
Government Spending	8.14	66.37	44.21	40.33
Expenditure/Revenue	22.53	9.05	12.15	18.05
Narrow Money	16.84	15.81	21.89	10.48
Broad Money	20.32	22.82	17.31	19.00
Real Wage	12.44	5.72	12.30	15.13

Table 4.9: Synchronization of the Latin American Cycles

The table reports the Harding-Pagan concordance statistic. Values close to one show evidence of a stronger degree of synchronicity.

	1878-1929				1930-1970			
	Arg	Bra	Chi	Mex	Arg	Bra	Chi	Mex
Arg	1.00	0.48	0.54	0.62	1.00	0.54	0.54	0.49
Bra	-	1.00	0.79	0.71	-	1.00	0.66	0.46
Chi	-	-	1.00	0.77	-	-	1.00	0.61
Mex	-	-	-	1.00	-	-	-	1.00
	1971-1987				1988-2004			
	Arg	Bra	Chi	Mex	Arg	Bra	Chi	Mex
Arg	1.00	0.59	0.53	0.76	1.00	0.41	0.82	0.71
Bra	-	1.00	0.71	0.59	-	1.00	0.47	0.47
Chi	-	-	1.00	0.41	-	-	1.00	0.65
Mex	-	-	-	1.00	-	-	-	1.00

Figure 4.1: Common Factors

Comparison of the first two common factors (F1, F2) extracted from series detrended with the Hodrick-Prescott (HP) or the Baxter-King (BK) filters.

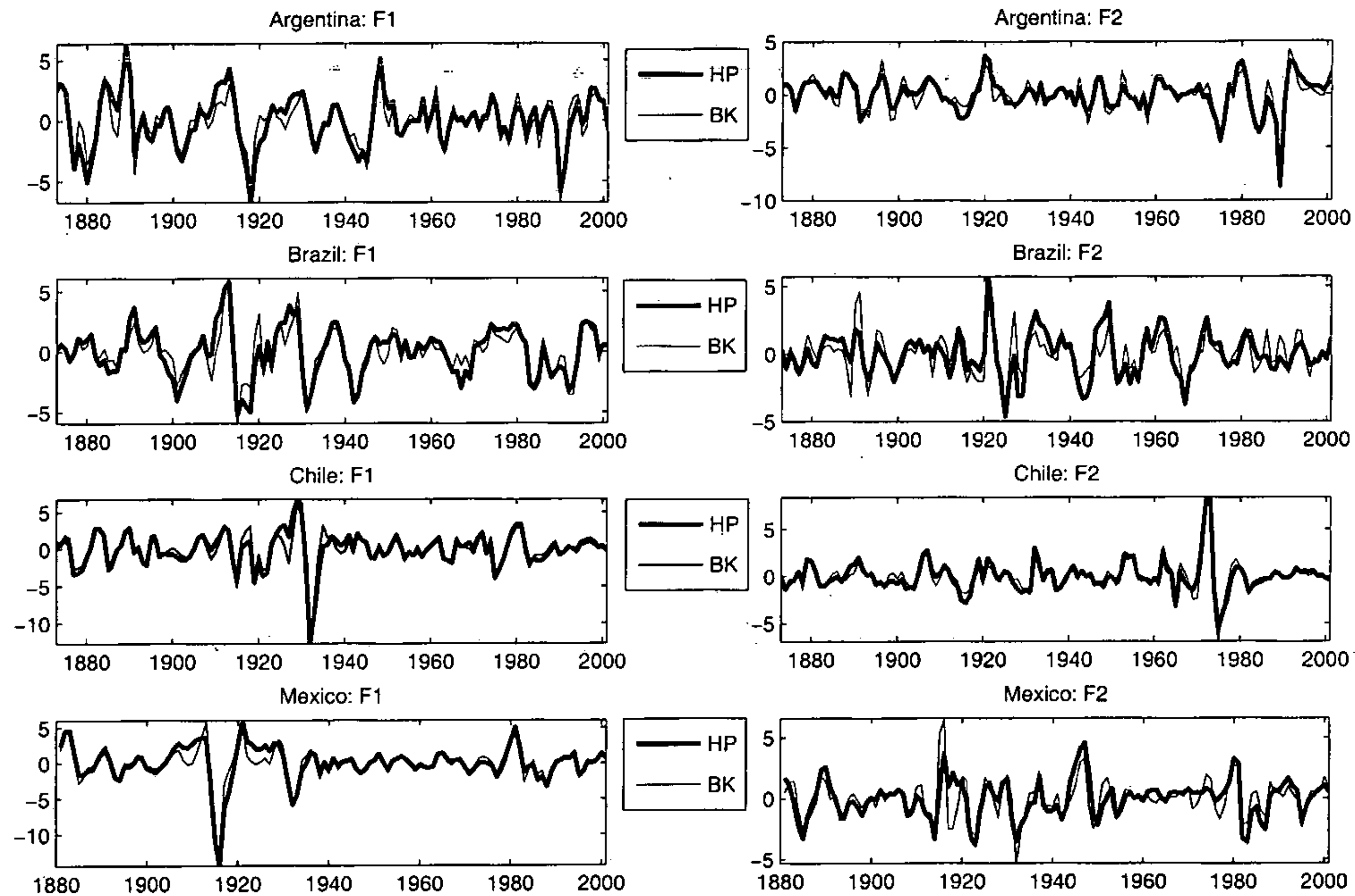
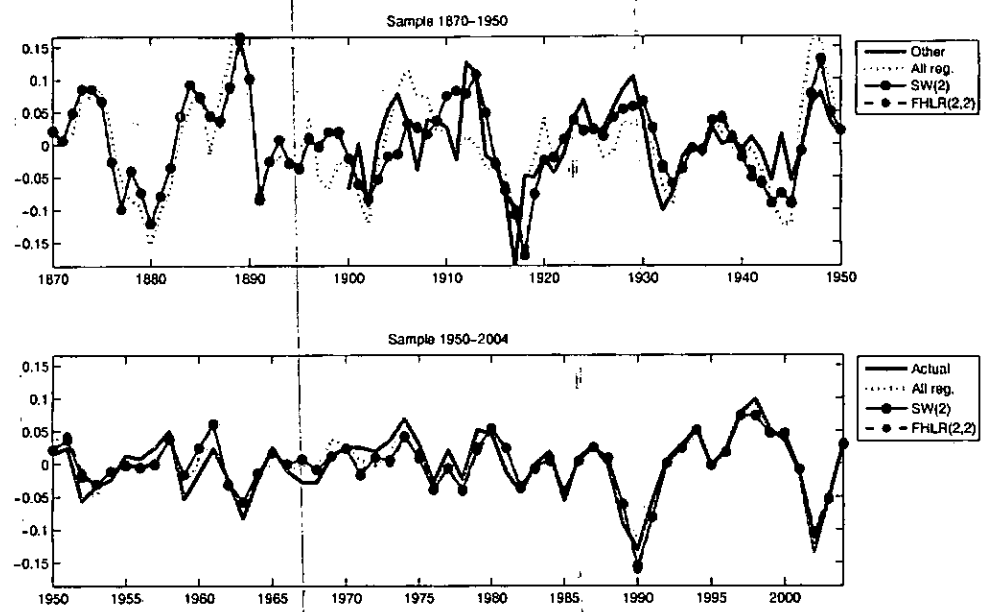


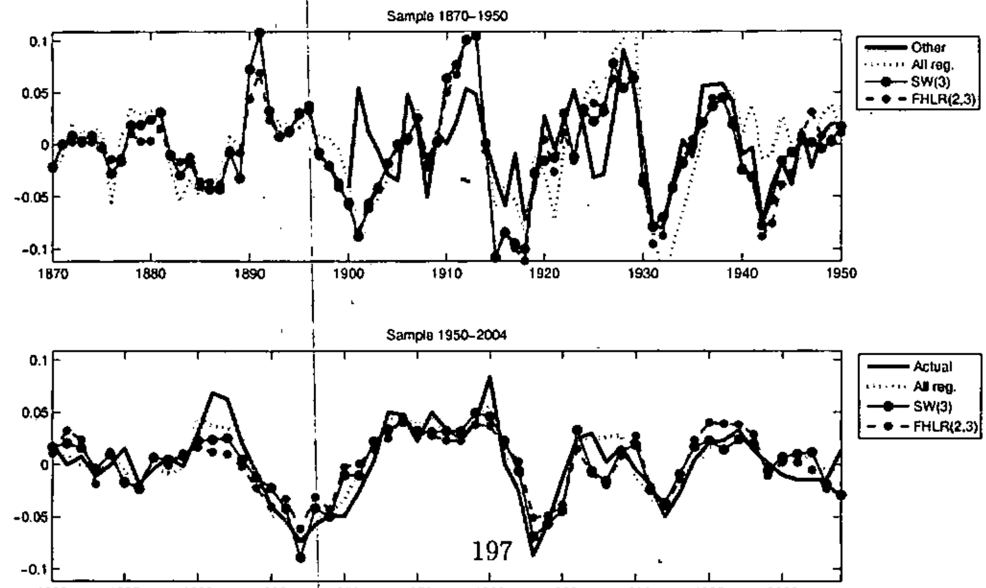
Figure 4.2: Actual and Backcasted Values of Cyclical Growth

Comparison between backcasted values, actuals, and other researchers' estimates of actual (Other) for Argentina, Brazil, Chile, and Mexico. All reg. denotes backcasted values that use all available regressors. The backcasting equation estimates the factors using either the Stock and Watson (2002) approach with r static factors (SW(r)) or the Forni et al (2000) dynamic factor approach with q, r dynamic and static factors, respectively (FHLR(q,r)). Each backcasting equation includes a constant term. The backcasting sample is 1870-1950 for all countries except Mexico (1878-1950).

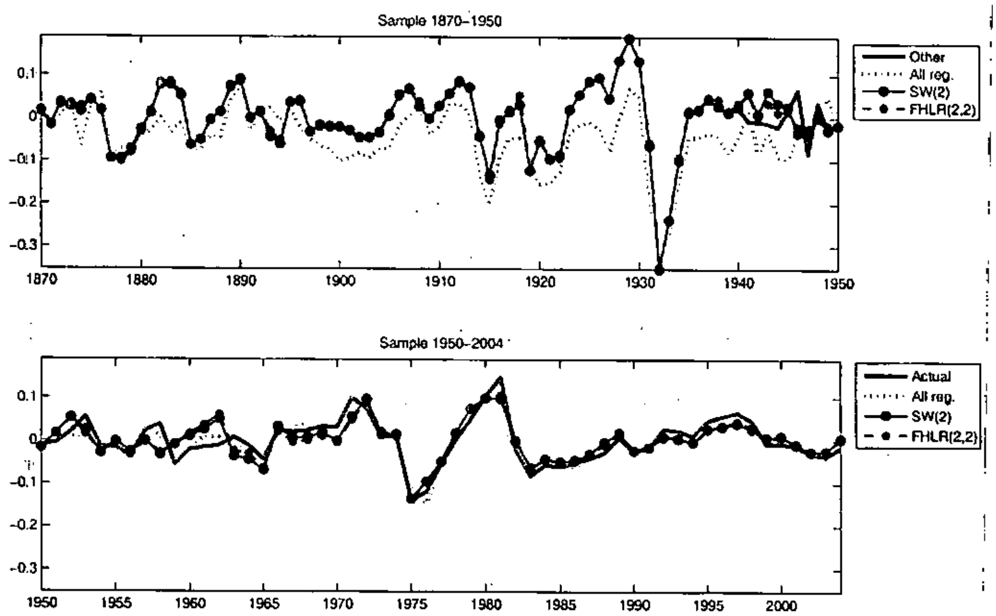
Panel A: Argentina



Panel B: Brazil



Panel C: Chile



Panel D: Mexico

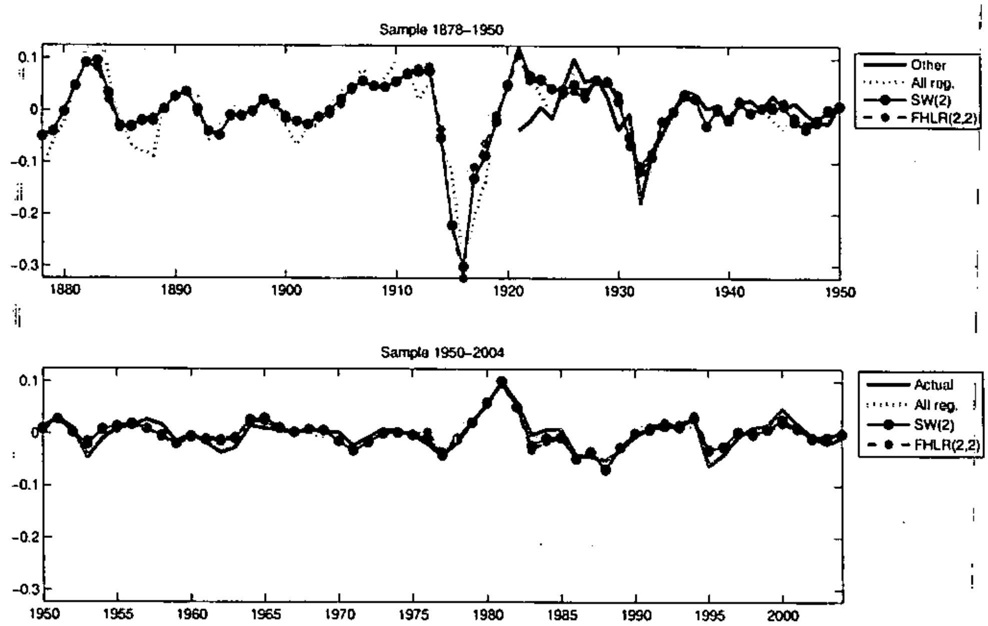
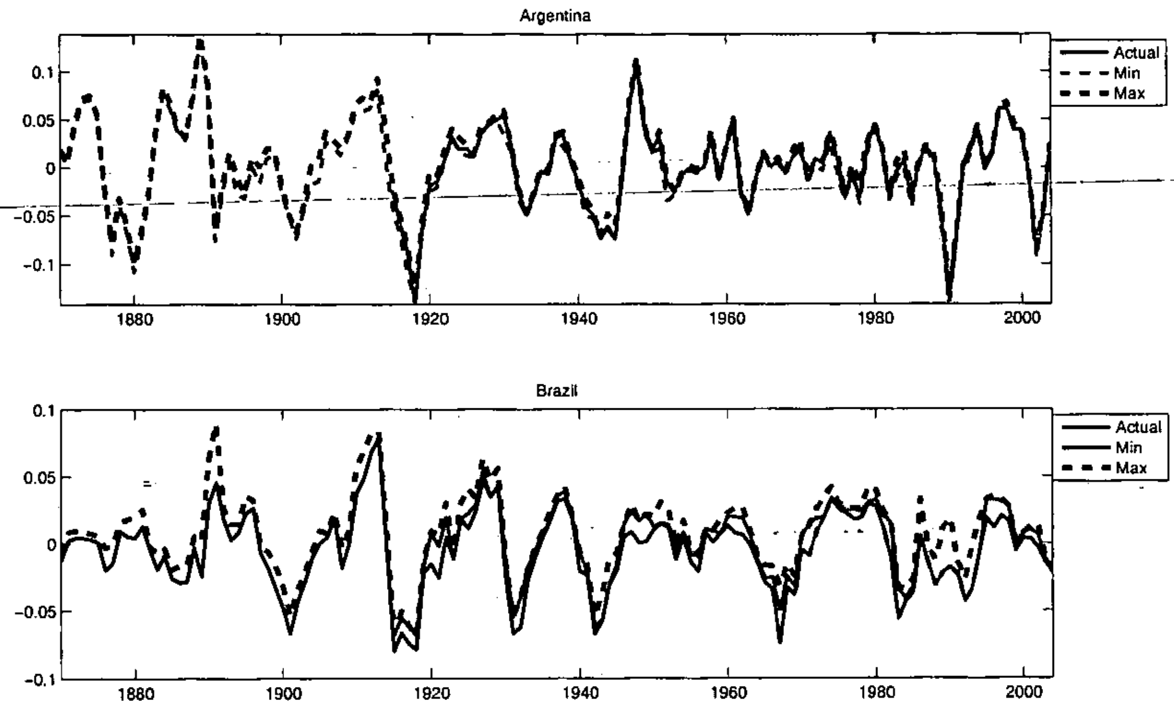


Figure 4.3: Robustness Check

Comparison of actual and backcasted values of business cycle. For each year the figures show the minimum and maximum backcasted value of cyclical output growth across models estimated using different numbers of factors and different data samples to estimate the backcasting equation or to extract common factors.



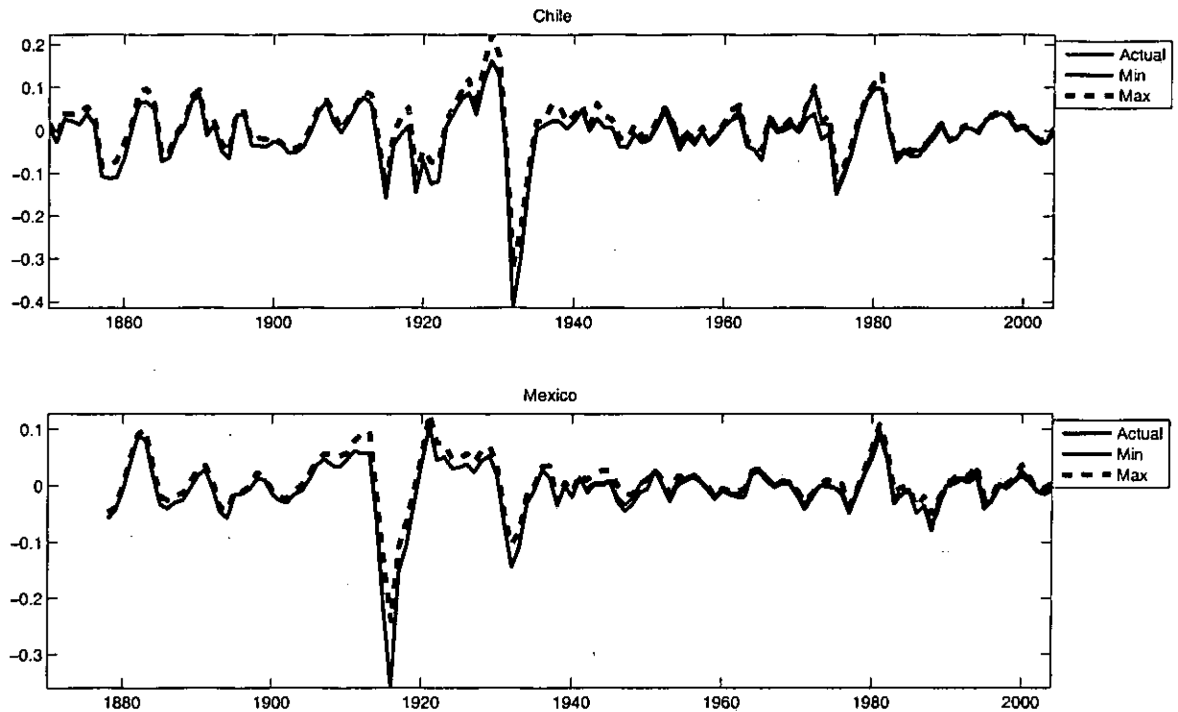


Figure continued from previous page.

Figure 4.4: Recursive Parameter Estimates

Recursive estimates of the coefficients used to obtain business cycle estimates from the backcasting equation. The first sample is 1960-2004, the last sample is 1920-2004, except for Chile (1940-2004). c is the intercept while b1,b2, and b3 are the slope coefficients for the first three common factors.

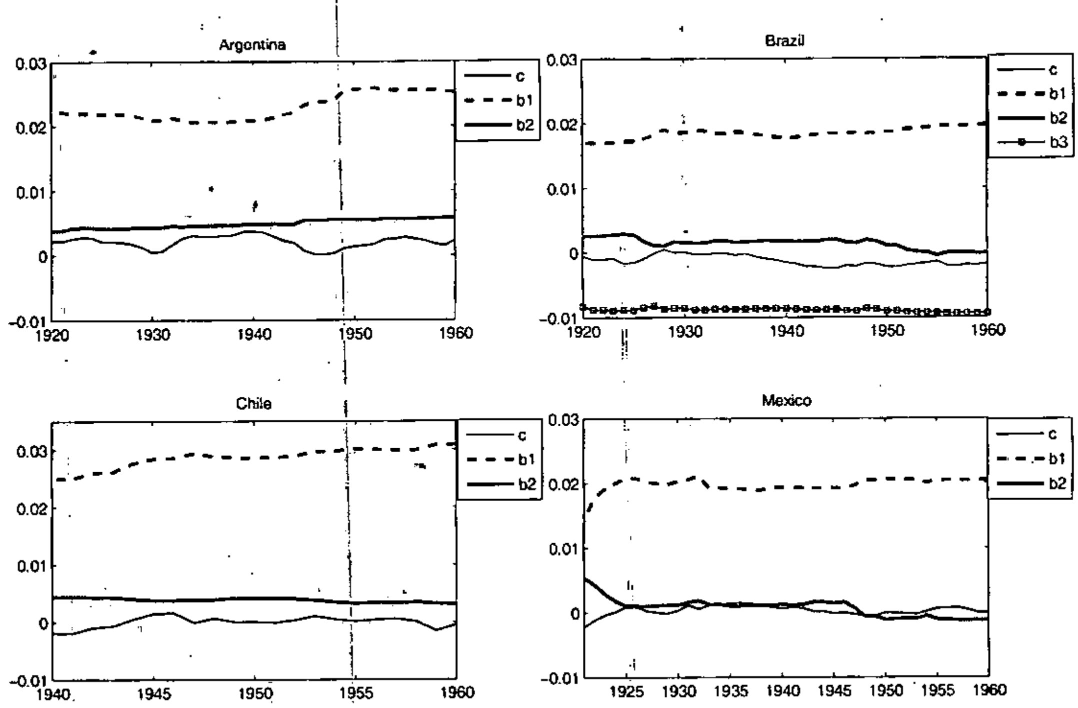


Figure 4.5: Estimates of the First Common Factor

Comparison of the first common factor constructed either on the basis of the full data set or solely on the basis of the sectoral output variables.

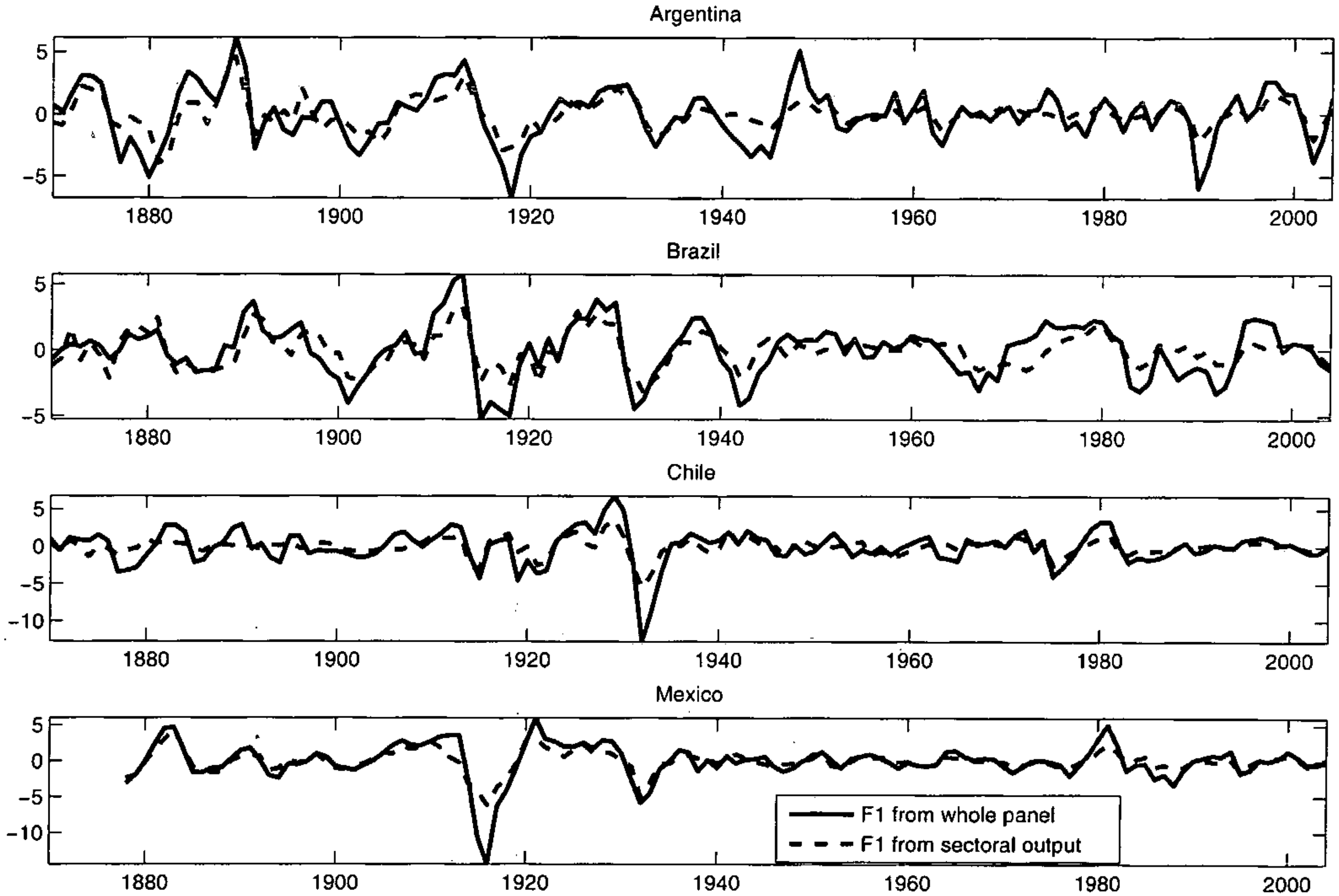


Figure 4.6: Actual and Backcasted Values of the US Business Cycle

Comparison of actual and backcasted values of business cycle for the US. Actual values are based on the Balke and Gordon series. All reg. denotes backcasted values that use all available regressors. The backcasting equation estimates the factors using either the Stock and Watson (2002) approach with r static factors (SW(r)) or the Forni et al (2000) dynamic factor approach with q, r dynamic and static factors, respectively (FHLR(q,r)). Each backcasting equation includes a constant term. The backcasting sample is 1870-1950.

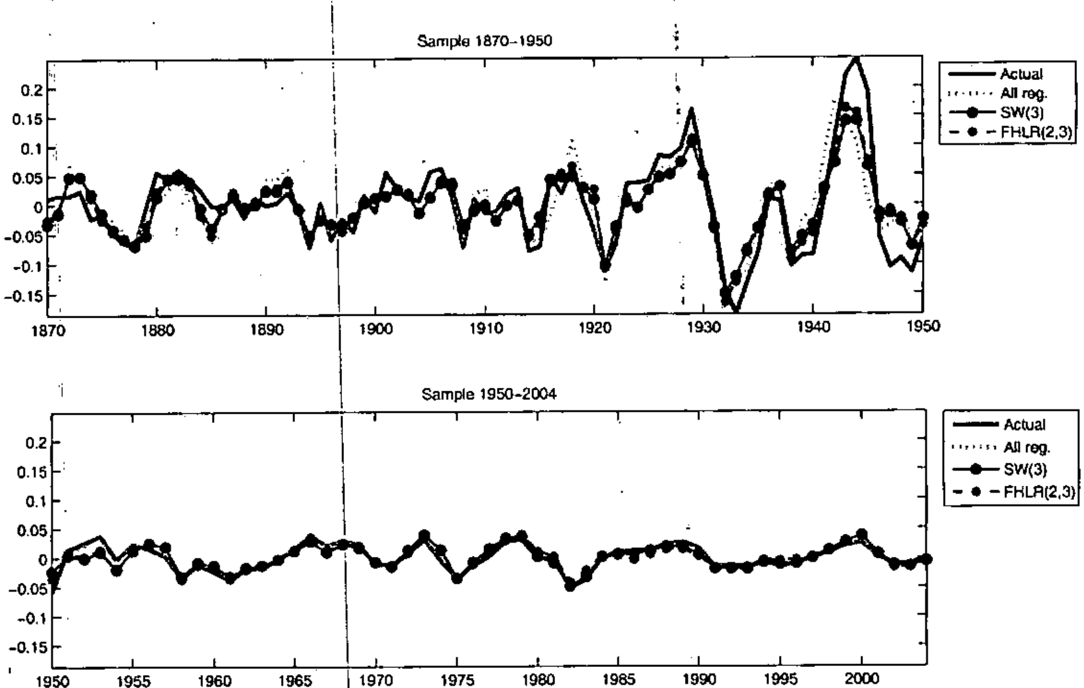
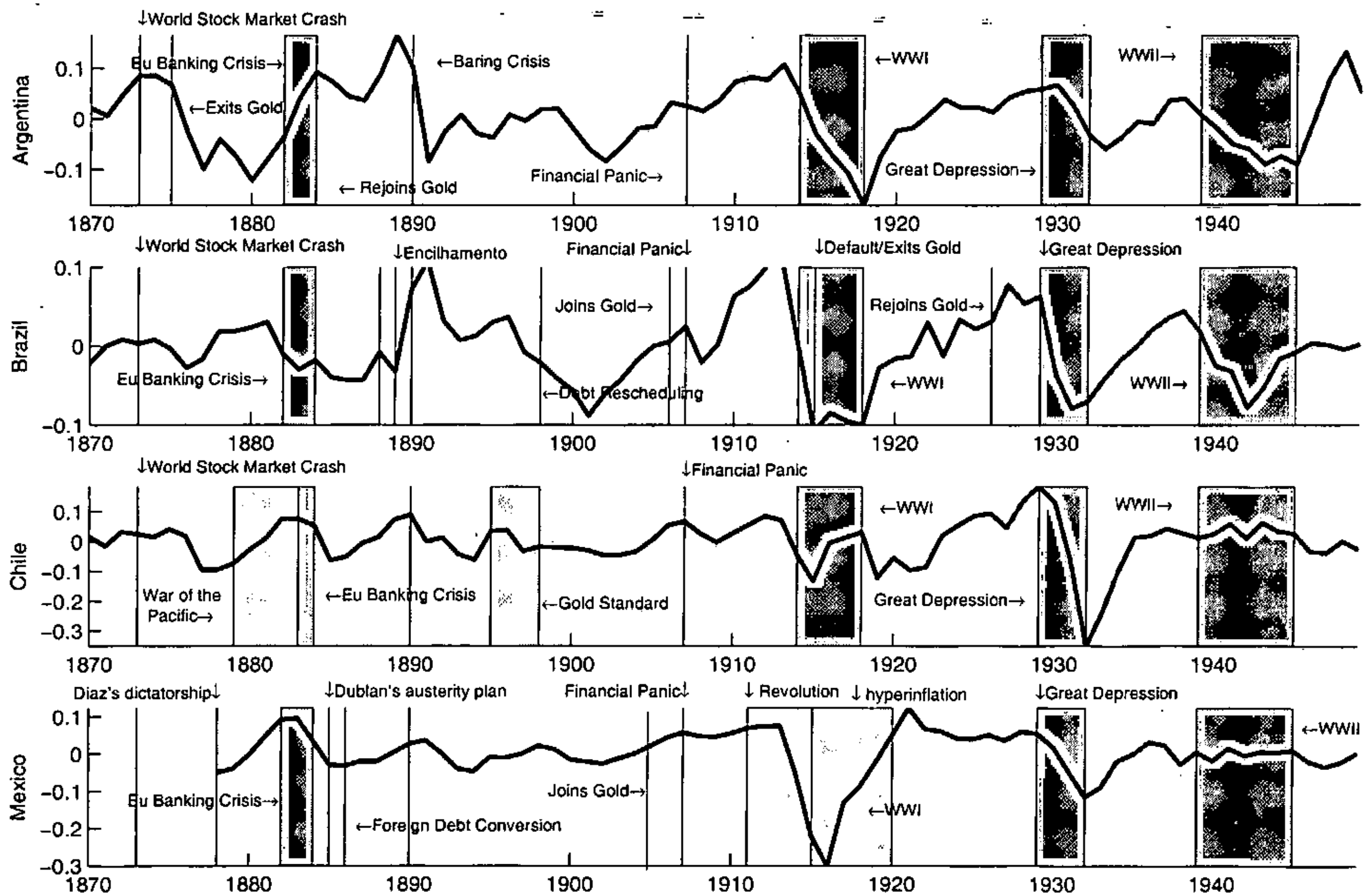


Figure 4.7: Historical Charts

Historical event charts. The panels report the backcasted business cycle for Argentina, Brazil, Chile and Mexico against world economic events (dark shaded area) and country-specific events (light shaded area).

Panel A: 1870-1949



Panel B: 1950-2004

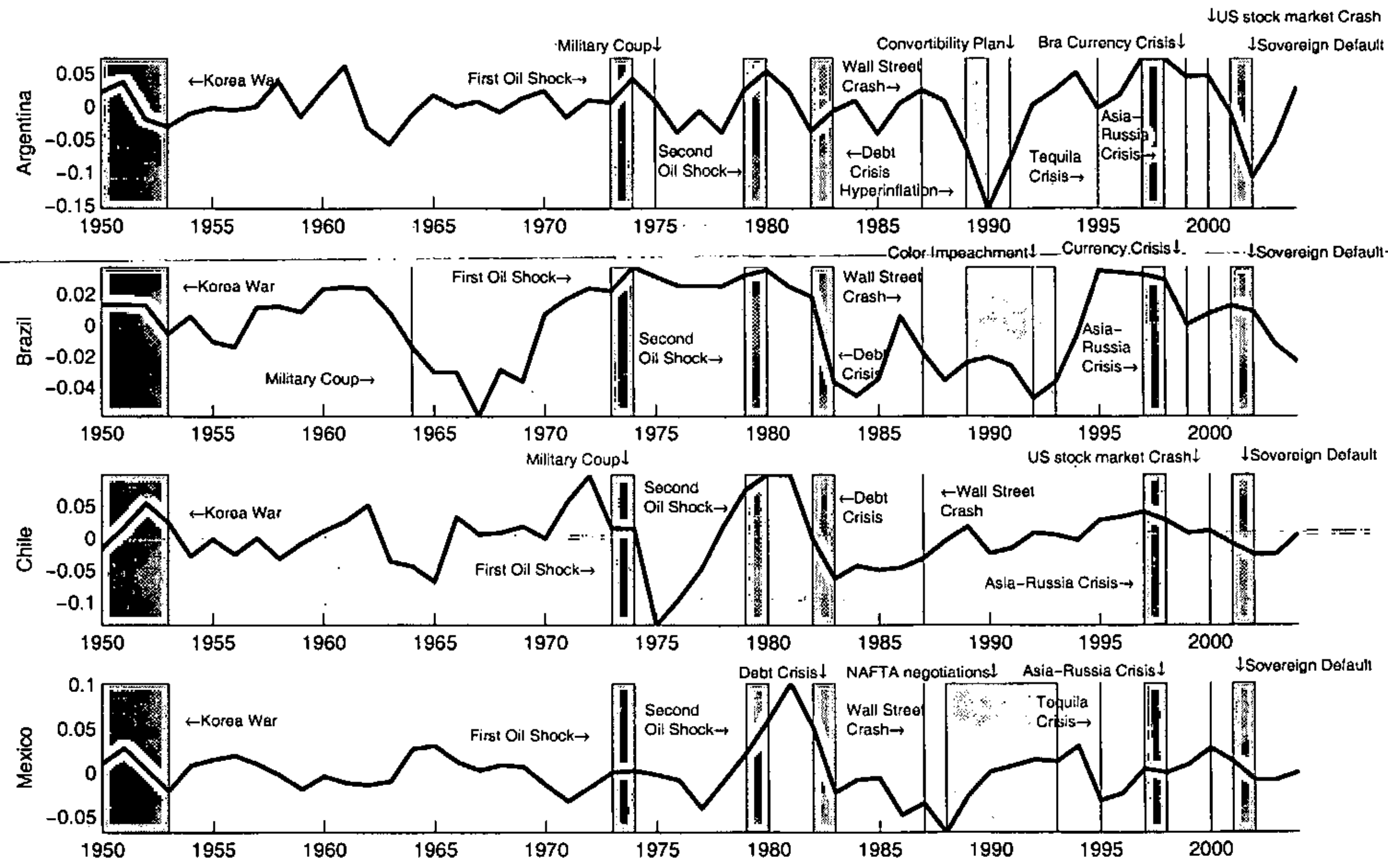


Figure continued from previous page.

Figure 4.8: Volatility

The figure reports 10-Year rolling window estimates of standard deviations of the backcasted business cycle. The top panel reports estimates for Argentina, Brazil, Chile, and Mexico. The backcasted values are based on a model using two common static factors (three for Brazil). The bottom panel reports estimates for the LA-4 region.