

学校编码: 10384

分类号_____ 密级_____

学号: 19020131152674

UDC_____

廈門大學

硕士学位论文

基于 Markov Chain Monte Carlo 方法对我
国股票市场收益率的预测

China's Stock Return Rate Prediction Based
on MCMC Method

刘若佳

指导教师姓名: 谭 忠 教 授

专业名称: 应 用 数 学

论文提交日期: 2016 年 4 月

论文答辩日期: 2016 年 5 月

学位授予日期:

答辩委员会主席: _____

评 阅 人: _____

2016 年 5 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（ ） 1.经厦门大学保密委员会审查核定的保密学位论文，于 年 月 日解密，解密后适用上述授权。

（ ） 2.不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

中文摘要

近年来, 股票慢慢成为金融市场中主要的理财工具, 对股票市场收益率的良好把握有助于人们优化手中的资产配置, 因此如何正确预测股票市场的收益率渐渐成为了经济金融研究中的热点问题. 为此, 一些学者开始试图应用先前的理论将股票收益率与其他变量相联系, 通过统计学方法并建立数学模型来寻找股票收益率与其他变量的数量关系, 尝试得到股票市场收益率的预测.

本文实证研究运用 2006 年 3 月至 2015 年 6 月间的月度数据, 通过我国的劳动收入增长率、消费增长率和国债利率三个因子对我国股票市场收益率进行预测. 基于传统的回归方程预测的种种弊端, 通过整理 VAR 模型的逻辑结构, 设置了各参数的先验分布, 依据贝叶斯理论对模型中各参数的后验分布进行推断. 并利用 MCMC 方法、Gibbs 抽样算法和 Kalman 滤波器理论对 VAR 模型中的参数进行迭代抽样, 以精确地估计各参数值, 进而完成对市场收益率的预测.

本文主要的贡献点包括三点: (1) 完善和补充了传统的回归方程, 通过引入股票市场收益率的条件期望 μ_t , 建立了 VAR(1) 模型, 意图通过控制 r_t 和 μ_t 两个序列的残差间相关系数 ρ_{uw} 的先验分布的信息含量, 来解决传统回归方程预测时因子选择不当、拟合精度低、估计参数较多的问题. (2) 本文在对 μ_t 进行迭代抽样时, 用到了 Kalman 滤波器理论, 可以避免原始序列的噪声对预测的影响, 获得更精准的估测值. (3) 通过自身模型与相同预测因子下回归方程的 R^2 进行对比, 定义新的指标 R^2 比来判断模型的有效性. 最终, 通过预测收益率与真实收益率的对比显示, 相对于传统的回归模型, 我们的模型预测效果较好, 尤其在判断股票市场收益率涨或者跌的问题上尤为准确.

关键词: 股票市场收益率 MCMC方法 Kalman滤波器

Abstract

In recent years, the stock has gradually been one of the most popular finance tools in financial markets, and forecasting the China's stock return helps people a lot in optimizing asset allocation in their hands. So predicting the stock market return rate in advance has gradually been a hot issue in financial study. As a result, more and more scholars get started on searching for the relationship between stock market return rate and some relevant variables, applying statistical methods and building mathematical models.

This paper is empirical research of prediction on China's stock return rate, through the following three factors, labor income growth rates, consumption growth rates and Treasury rates. Our dataset is the monthly data from March, 2003 to June, 2015. This paper is based on the traditional regression function and added the conditional expected stock return rate μ_t to our VAR(1) model, and make an increase or decrease on the information content of the parameter ρ_{uw} 's prior distribution to avoid traditional regressions' shortages of inappropriate predictors selection, lower fitting degree, and the high number of estimated parameters' amount. And ρ_{uw} is the correlation coefficient between the two sequences r_t and μ_t 's residuals. Based on the analysis of statistical model, we make the prior distribution of some parameters as an assumption, applied Bayesian theory on getting the posterior distribution of the model parameters, using the method of MCMC, Gibbs sampling and Kalman filter to iteration sampling to sample VAR model's parameters, to accurately obtain each parameter's Markov stationary distribution. We make each parameter ten thousand samplings, and regard the mean of sequences after reaching the stationary distribution as the final estimate of the parameter, and finally get the prediction of China's stock return rate. In addition, the paper makes a variable R^2 ratio to compare

the R^2 of our model with the traditional regression. In the end, we are satisfied with the final prediction through comparing our predicted return rate to the real return rate, especially in judgment of stock return's rise and fall.

Key words: Stock Return Rate MCMC Method Kalman Filter

厦门大学博硕士论文摘要库

目 录

中文摘要	I
英文摘要	II
中文目录	IV
英文目录	VI
第1章 引言	1
1.1 研究背景	1
1.2 国内外研究综述	1
1.3 本文主要工作	6
第2章 预备知识	8
2.1 Kalman滤波	8
2.2 MCMC方法与Gibbs抽样	12
2.3 有关贝叶斯估计的正态分布和逆维希特分布	14
第3章 预测系统建立	15
3.1 VAR模型建立	15
3.2 参数先验分布和初值设定	18
3.3 MCMC方法-Gibbs抽样	20
第4章 中国股票市场收益率预测的实证分析	33
4.1 模型建立和数据说明	33
4.2 MCMC算法迭代抽样	34
4.3 模型的评价标准： R^2 比	37
4.4 预测结果	39

第5章 结论	42
参考文献	44
致谢	47

厦门大学博硕士论文摘要库

Contents

Chinese Abstract	I
English Abstract	II
Chinese Contents	IV
English Contents	VI
1 Introduction	1
1.1 Physical Background	1
1.2 Review of Research	1
1.3 Main Work	6
2 Preliminary Knowledge	8
2.1 Kalman Filter	8
2.2 MCMC Method and Gibbs Sampling	12
2.3 $N(\mu, \sigma^2)$ and $IW(\Psi, \nu)$	14
3 Building Prediction Frame	15
3.1 Building VAR(1) Model	15
3.2 Make Assumptions of Prior Distribution and Initial Value	18
3.3 MCMC-Gibbs Algorithm	20
4 Empirical Study of China's Stock Return Rate Prediction	33
4.1 Modeling and Data Explanation	33
4.2 MCMC Algorithm and Iteration Sampling	34

4.3 Evaluation Standard: R^2 Ratio	37
4.4 Result of Our Prediction	39
5 Conclusion	42
References	44
Acknowledgements	47

厦门大学博硕士学位论文摘要库

第1章 引言

1.1 研究背景

近年来, 股票慢慢成为金融市场中除储蓄和债券之外最主要的理财工具. 实际上, 相对于西方发达国家的金融市场而言, 我国股票市场的起步和发展是较晚的; 但在近几年经济飞速增长与发展的背景下, 我国的股票市场已趋于完善. 随着股票被越来越多的投资者所青睐, 对股票市场的收益率的预测越来越有助于人们优化手中的资产配置, 因此提前预测股票市场收益率也逐渐成为金融中的一个热点问题. 沪深 300 指数是反映中国沪深两个股票市场整体走势的跨市场指数, 它的出现切合了市场需求, 为人们对市场收益率的预测增加了一项用于估计市场收益率走势的参照, 且能取到数据的时间年限较长, 因此我们可以认为沪深 300 指数是可以一定程度上代表全市场股票收益状况的重要指标. 本文用沪深 300 指数的增长率代表我国股票市场收益率来作为预测变量. 对我国股票市场收益率预测的本质就是寻找和利用与股票收益率相关的若干个变量, 通过统计学方法并建立数学模型来寻找二者之间的联系, 最终得到对股票市场收益率的精确预测.

1.2 国内外研究综述

从 2000 年开始, 国外对股票市场收益率预测的研究就已经有了雏形. 迄今为止, 国外学者对这个问题的理论研究, 从简单的回归到最大似然估计、贝叶斯理论, 到由贝叶斯理论衍生出的各种算法, 一直在飞速的进步与发展.

2000 年, NICHOLAS BARBERIS^[1]利用股息率和国债利率, 通过回归方程建立了关于股票市场收益率 r_t 对 x_t 的回归模型:

$$r_{t+1} = \alpha + \beta'x_t + \epsilon_{t+1}$$

其中, x_t 为 t 时刻的股息率和国债利率组成的二维向量, r_t 为 t 时刻的股票市场的收益率. 如此, t 时刻的期望市场收益率 μ_t 为

$$\mu_t = \alpha + \beta' x_t.$$

2008 年, Lubos Páster 和 Robert F. Stambaugh^[2]运用最大似然估计方法通过股息率与国债率对美国的股票收益率进行了预测, 他们发现最大似然估计方法的收敛速度太过缓慢, 并且有些参数的似然估计参数不容易求解, 同时作者还指出传统的回归模型存在着两点不足, 首先, 回归模型对因子与预测变量的线性关系要求太过苛刻, 股票预期收益率很可能和历史收益率(滞后项)相关, 如此模型很可能由于不满足线性关系而导致预测精度降低; 其次, 一般情况下回归方程的残差项 ϵ_t 都受历史收益率 r_t 影响, 会出现 $Cov(\epsilon_t, \epsilon_{t+1})$ 不为 0 而导致 $\mu_t = \alpha + \beta' x_t$ 不成立的情况, 文中将这类因子叫做“不完美因子”. “不完美因子”会给回归模型带来虚假的预测信息, 因此促使我们不得不寻找其他模型来解决这个问题, 于是他们对股票市场的条件预期收益率 μ_t 单独建立了 AR(1) 模型, 再联立 r_t , 可以得到 VAR(1) 模型, 即

$$\begin{cases} r_t = \mu_{t-1} + u_t, \\ \mu_t = (1 - \beta)E_r + \beta\mu_{t-1} + w_t. \end{cases} \quad (1-1)$$

其中, $\mu_{t-1} = E(r_t | D_{t-1})$ 为 t 时刻的条件期望收益率, 为不可观测的变量; u_t 为 t 时刻的非预期收益, $E(u_t | D_{t-1}) = 0$, $D_{t-1} = \{r_1, \dots, r_{t-1}\}$.

其实早在 2003 年, WAYNE、SERGEI 和 TIMOTHY^[3]就针对上述“不完美因子”提出过相似的概念, 文中明确指出: 对于正常的预测因子, 一定有 $\rho_{uw} < 0$. 当预测因子的 $\rho_{uw} \geq 0$ 时, 在预测模型中加入该因子便会降低预测的效果. 于是我们将问题转化为: 如何通过 μ_t 来处理“不完美因子” x_t , 并最终对 r_t 完成准确的预测.

现在我们将预测因子 x_t 考虑其中, 假设 x_t 满足 AR(1) 模型, 即

$$x_t = Ax_{t-1} + (I - A)E_x + v_t \quad (1-2)$$

将 (1-1) 和 (1-2) 联立组成新的 VAR 模型. 值得注意的是, 我们通过 ρ_{uw} 将 μ_t 和非

预期收益 u_{t+1} 建立了联系; 由于

$$\rho_{x\mu} = \rho_{vw} \left[\frac{(1 - \beta^2)(1 - A^2)}{(1 - \beta A)^2} \right]^{0.5}$$

则 x_t 完全通过 v_t 和 w_t 来影响 μ_t , 进而影响 r_t . 当 x_t 是“不完美因子”时, 我们可能面临两个问题: (一) 历史收益率也受到预测因子的历史数据 (滞后项) 的影响; (二) $\rho_{uw} < 0$ 不满足. 上述两个问题都会导致回归模型失效.

于是, 我们试图通过建立 r_t, x_t, μ_t 的 VAR 模型, 并通过随机模拟生成模型参数以及残差的 Markov 链, 令 r_t, x_t, μ_t 之间只通过残差项相互影响, 由此就解决了问题一的模型非线性问题. 对于问题二, Lubos Páster 和 Robert F. Stambaugh^[4] 在另一篇文章中指出: 可以尝试控制 ρ_{uw} 的信息含量来解决. 有关于 ρ_{uw} 的重要性, 文中有如下陈述:

$$\begin{aligned} E(r_{t+1}|D_t) &= E_r + \sum_{s=1}^t \beta^{t-s} M_s \varepsilon_s + \sum_{s=1}^t \beta^{t-s} N_s v_s \\ M_s &\rightarrow M \\ M &= (\beta Q + Cov(u, w|v))(Q + Var(u|v))^{-1} \end{aligned}$$

ρ_{uw} 直接决定 M 的值进而影响对 $E(r_{t+1}|D_t)$ 的预测; 而针对问题二, 我们尝试通过改变 ρ_{uw} 来削弱 x_t 对 r_t 的影响. 具体做法如下: 设

$$\rho_{vw} = \rho_{uv}\rho_{uw} + \rho_{\Delta}, \text{ 其中 } \rho_{\Delta}^2 \leq (1 - \rho_{uv}^2\rho_{uw}^2)$$

当我们给 ρ_{uw} 的先验分布增加信息含量时, 由于 x_t 是不完美因子, 当 ρ_{uv} 接近于 ± 1 时, 说明 x_t 对 μ_t 有较大影响, 这种情况我们不是很希望见到. 基于 ρ_{uv} 接近于 ± 1 , 因而 ρ_{Δ} 很小, 我们可以忽略不计, 则有 $\rho_{vw} \approx \rho_{uv}\rho_{uw}$. 如此变量 ρ_{uw} 便将 u_t, v_t, w_t 联系起来, 他们之间就可以相互影响. 具体 ρ_{uw} 和 ρ_{vw} 的关系如图 1.1 和图 1.2 所示,

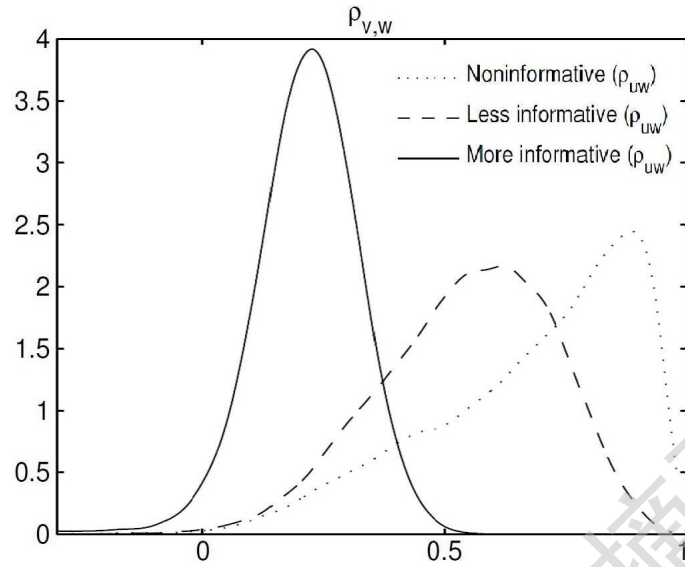


图 1.1 不完美因子的 ρ_{uw} 对 ρ_{vw} 的影响

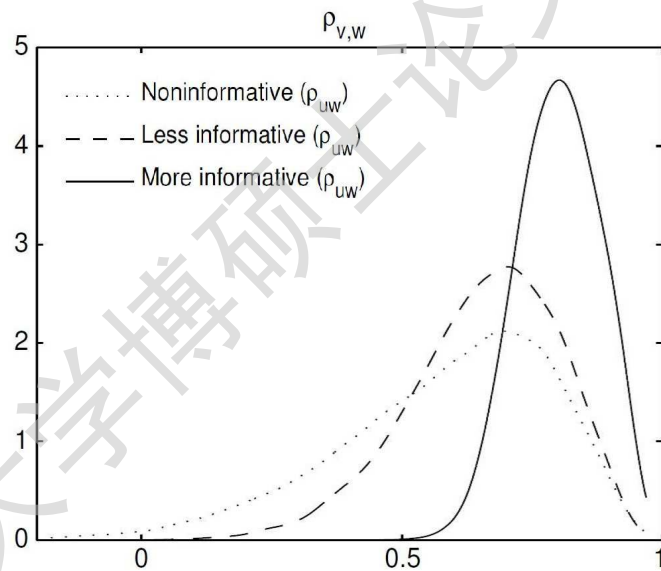


图 1.2 完美因子的 ρ_{uw} 对 ρ_{vw} 的影响

如上图, 当逐步加大 ρ_{uw} 的先验分布信息含量时, ρ_{vw} 会降低, 因此 x_t 和 μ_t 的相关系数 $\rho_{x\mu}$ 也会降低; 而对于完美因子而言, 增加 ρ_{uw} 的先验分布信息含量会使 ρ_{vw} 增大, 同样有益于模型的预测. 因此增大 ρ_{uw} 先验分布信息含量对我们模型的预测效果有很大的益处. 于是我们决定通过增加 ρ_{uw} 的先验分布信息含量来降低“不完美因子”对模型中 r_t 未来值预测的影响, 进而改善预测的精度; 考虑到这点, 本文建立了 r_t, x_t, u_t 的 VAR 模型, 通过给 ρ_{uw} 赋予含信息较多的先验分布, 解决含有“不完美因子”的预测问题.

如此, 模型建立本身的问题虽然解决了, 但数据上的问题却仍然困扰着我们. 由于中国股票市场的数据量有限且存在异常值, 我们不得不采用一些特殊的方法来解决数据数量和质量上的缺陷.

近年来, 为提高模型信息和样本信息的利用率, 避免参数估计数量过多以及可观测的样本数量有限等问题, 国内越来越多研究者去尝试应用贝叶斯理论来解决模型参数的估计的问题. 他们以贝叶斯估计为理论根基, 将建立的 VAR 模型参数和潜在的波动变量(残差)都当作随机变量处理, 通过假设先验分布和似然函数来计算参数的边缘后验分布来完成对参数的估计以及对变量的预测. 但在大多数情况下参数和变量的联合分布以及后验分布都是非常不易计算的高维分布, 传统的 Bayesian 估计方法无法求取解析解, 以至于不能抽取样本估计参数并用来做预测.

MCMC 方法 (Markov Chain Monte Carlo) 的出现恰好解决了我们的难题. MCMC 方法主要致力于是估值与抽样, 它主要是对传统随机模拟的理论更新, 它不仅克服了传统方法还未解决的高维动态模拟的问题, 并且间接地从参数的联合分布中生成随机样本, 避免了一一假设参数的后验分布的过程. 从这个层面上理解, 它提高了传统 Monte Carlo 模拟的精度与效率. 只要模拟次数足够多, 我们就可以得到收敛到精确值附近的样本统计量. 由 Markov 理论, 经过充分长时间, 待抽样参数的后验分布会收敛到一个平稳分布, 我们可以以这个结果作为抽样依据估测出所有参数. 因此 MCMC 方法的关键就变成了, 如何建立一个平稳分布的 Markov 链. MCMC 方法中通常应用 Metropolis-Hastings 算法或者 Gibbs 算法抽样生成该 Markov 链, 即先通过先验分布和似然函数的假设得到各参数的后验分布, 之后依据后验分布对其抽样, 然后用 Monte Carlo 方法在已抽出的样本中做所需的统计推断, 获得各参数的 Markov 链, 进而获得全部参数的估计值, 对我们想要的变量值做出预测. 另外两点需要说明: 第一, 时间序列中样本序列异常值的存在经常会对模型参数的估计有所影响, 进而降低模型的预测精度, 而 Gibbs 算法抽样对参数信息的实时更新能起到减少误差的产生和累积的作用. 第二, 进行状态变量的更新时经常用到的 Kalman 滤波方法, 必须保证状态空间模型的残差项服从正态分布.

国内已经有很多应用 MCMC 方法的解决问题的实例. 薛文骏(2013)^[5]通过贝叶斯分位数回归建立模型并使用 MCMC 方法和 Gibbs 抽样对我国投资者对于市场信息是否存在过度反应和反应不足进行了实证分析. 邓慧敏(2014)^[6]中基于贝叶斯理论运用 MCMC 方法和二次 Gibbs 抽样搭建了 VAR 模型探讨股票回报率、通胀率波动和通胀率趋势三个变量之间的相互作用关系, 最终得到了 MCMC 方法应用在 VAR 模型中能更准确地模拟出变量之间非线性波动的动态特征. 朱钧钧和谢识予(2011)^[7]通过 MCMC 方法来优化 MS-TGARCH 模型来研究我国股票市场波动率不对称的特征.

目前为止, 国内的实证研究还尚未有用应用 MCMC 方法和 Gibbs 抽样建立 VAR 模型对中国股票市场收益率进行直接预测的实例. 上文中已经阐述的 NICHOLAS BARBERIS(2000)^[1]指出国债利率为“不完美因子”, 且其是影响市场收益率的因素, 还有 MARTIN LETTAU 和 SYDNEY LUDVIGSON(2001)^[8]指出股票市场的收益率与劳动收入增长率和消费增长率显著有关, 因此本文就综合考虑以上三个因素, 建立新的预测模型来探讨利用我国的劳动收入增长率、消费增长率和国债利率这三个因子来预测我国股票市场收益率.

1.3 本文主要工作

本文基于贝叶斯理论搭建了 VAR 模型, 并通过我国劳动收入增长率、消费增长率和国债利率三个变量对中国股票市场收益率进行了预测, 并解决了传统回归统计方法中忽略“不完美因子”和变量间非线性关系的不足. 正文先假设了各参数的先验分布、似然函数, 并依据贝叶斯理论推出了各参数的后验分布. 之后运用了 MCMC 方法获得各参数的平稳分布, 并用 Gibbs 算法抽样对模型中各参数和变量进行迭代抽样, 其中对某些参数变量的抽样过程中运用了 Kalman 滤波器理论来获得更精确的参数预测. 如此, 最终得到了对我国股票市场收益率的预测. 本文的主要内容如下:

第一章为本文的引言部分, 介绍了本文的研究背景、国内外文献综述以及本文用到的模型简介.

Degree papers are in the “[Xiamen University Electronic Theses and Dissertations Database](#)”.

Fulltexts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.