

样本外推效果与改进方法

研究目的与结论

自从发布《宏观经济变量与股市关系系列报告》之后，我们提出的宏观经济变量对股市的影响模式的研究框架受到了很多投资者的广泛关注。同时，我们感到仍旧有很多方面需要继续探索，尤其是领先多因素模型的样本外效果究竟如何，以及如何改进预测效果，这篇报告就是试图解答这两方面的问题。

我们发现宏观经济变量对股市的预测效果是明显的，并且同时也发现有两个方面的因素可以提高宏观经济变量对股市的预测效果：

第一、及时纳入新的领先变量可以提高对股市的预测效果。我们使用 5 年的数据建立领先多因素模型，分别逐年、逐月外推对股市进行样本外预测，发现逐月外推相比逐年外推可以提高预测的准确性，样本外的预测方向正确的概率达到 58%。

第二、对宏观经济变量的趋势具有前瞻性可以提高对股市的预测效果。我们在对宏观经济变量的趋势具有前瞻性的情况下建立多因素模型，逐月外推对股市进行样本外预测，发现对宏观经济变量的趋势具有前瞻性的多因素模型可以提高预测的准确性，样本外的预测方向正确的概率达到 61%。

外推效果与改进之后的测试结果

	一年更新一次模型		一个月更新一次模型	
识别状态时 有无前瞻性	无前瞻性	有前瞻性	无前瞻性	有前瞻性
coefficient	0.01	0.30	0.48	0.54
t-test	0.10	2.31	3.23	3.94
Rsquare	0.00	0.04	0.08	0.11

2011-01-30

领先多因素模型样本外择时效果图



领先多因素模型样本外择时效果对比图



领先多因素模型样本外择时效果对比图



相关研究（点击下载）：

[《宏观经济变量与股市关系系列报告》](#)

部门微博：

<http://weibo.com/cjscfe>

分析师：

范辛亭

(8621) 68751859 (fanxt@cjsc.com.cn)

执业证书编号：S0490510120008

联系人：

范辛亭

(8621) 68751859 (fanxt@cjsc.com.cn)

正文目录

一、 探求多因素模型的样本外效果.....	3
二、 多因素模型外推效果及改进方法.....	3
1. 多因素模型外推效果	3
2. 及时纳入新的领先变量对预测效果的改进.....	5
3. 前瞻性对预测效果的改进.....	8
三、 总结与思考.....	11

图表目录

图 1: 领先多因素模型样本外择时效果图 (2001.1-2011.10)	5
图 2: PPI 与上证综指的关系图 (1996.12-2001.11)	7
图 3: PPI 与上证综指的关系图 (2006.1-2010.12)	7
图 4: 领先多因素模型样本外择时效果图 (2001.1-2011.10)	8
图 5: M1 同比增速及其状态, 有前瞻性与无前瞻性对比	9
图 6: 多因素模型外推一年的择时效果	11

表 4: 领先多因素模型 (样本区间 1996.1-2000.12、2000.1-2004.12)	4
表 5: 领先多因素模型 (样本区间 2001.1-2005.12)	4
表 6: 领先多因素模型 (样本区间 2002.1-2006.12、2006.1-2010.12)	4
表 7: 领先多因素模型 (样本区间 2005.10-2010.9、2006.1-2010.12)	6
表 8: 领先多因素模型 (样本区间 1996.12-2001.11)	6
表 9: 样本外预测效果检验结果对比	8
表 10: 无前瞻性与有前瞻性识别趋势建立的多因素模型对比 (样本区间为 2000.1-2004.12)	10
表 11: 无前瞻性与有前瞻性识别趋势建立的多因素模型对比 (样本区间为 2006.1-2010.12)	10
表 12: 样本外预测效果对比	10

一、探求多因素模型的样本外效果

自从发布《宏观经济变量与股市关系系列报告》之后，我们提出的宏观经济变量对股市的影响模式的研究框架受到了很多投资者的广泛关注，高兴之余，深深地感受到我们所接触的问题的复杂性，也时常感到某些方面还值得进一步的探索。《宏观经济变量与股市的关系系列报告》帮助我们梳理了许多宏观经济变量与股市的关系，我们时常关注的宏观经济指标究竟与股市是什么样的关系，是长期或是短期，正向或是反向，领先或是滞后？我们在上述系列报告里面回答了这个问题。

一年来我们坚持用系列报告中建立的多因素模型对股市的涨跌做出预判并与大家分享，样本外领先变量对股市的预判效果虽然不错，但是仍然难以达到样本内的解释能力，那么，领先变量在样本外对股市的预判效果究竟如何，我们对股市的预测收益率与股市的实际收益率是否显著相关，关系多大，宏观经济变量与股市的领先滞后关系是否稳定，我们是不是需要根据最新的数据对二者关系进行调整，都是我们十分关心的问题。

这篇报告就是从这几个角度出发，与大家分享我们近期的一些研究成果。这篇报告首先介绍了多因素模型的滚动外推效果，接下来介绍了几个改进外推效果的方法。

二、多因素模型外推效果及改进方法

1. 多因素模型外推效果

为了考察多变量的回归结果是否对实际的预测有帮助，我们将多变量回归结果进行滚动外推对股市的收益率进行预测，并且着重研究了样本外的预测效果，对样本外的预测结果进行了检验。

我们采取滚动外推的做法对样本外的股市进行预测，即：从 1996 年 1 月开始，选取 5 年的数据（1996 年 1 月至 2000 年 12 月，共 60 个样本点）检验宏观经济变量与股市的关系，并找出领先的变量建立多因素模型，并且对未来 12 个月，即：2001 年 1 月至 2001 年 12 月的股市收益率进行预判，接下来，加入这 12 个月新的宏观经济变量的数据，仍旧保持用最新 5 年的数据建立新的领先多因素模型，再对未来 12 个月的股市收益率进行预判，以此方法逐步外推。这样，我们在 2000 年至 2010 年这 10 年每年年底都会建立一个领先多因素模型，并且根据模型预测了 2001 年 1 月至 2011 年 10 月每个月的股市收益率。

为了检验领先多因素模型对股市的预测效果，使用下面的模型：

$$r_t = \beta_0 + \beta_1 r_{t-1} + \varepsilon_t \quad (3)$$

其中， r_t 是上证综指的实际月度收益率， r_{t-1} 是领先多因素模型的预测收益率，通过考察系数 β_1 的显著性来判断多因素模型的样本外预测效果。

下面几个表格列举了根据不同时间段内宏观经济变量与股市的关系建立的领先多因素模型：

表 1: 领先多因素模型（样本区间 1996.1-2000.12、2000.1-2004.12）

	样本区间：1996.1-2000.12			样本区间：2000.1-2004.12		
	常数项	M0 同比增速	M2 同比增速	常数项	工业增加 值同比	M0 同比增速
j		6	9		4	1
k		-2	-2		-2	-1
coefficient	-0.94	4.71	11.16	-1.66	4.25	-2.70
t-test	-0.67	2.34	3.38	-1.48	3.20	-2.05
Adj-Rsquare	0.25			0.16		

表 2: 领先多因素模型（样本区间 2001.1-2005.12）

样本区间： 2001.1-2005.12	常数项	工业增加值 同比	发电量同比	固定资产投资 来源合计 累计同比	美元指数
j		2	4	1	4
k		-2	-2	-2	-2
coefficient	0.19	3.30	-3.37	4.64	-3.98
t-test	0.14	2.62	-2.67	3.49	-2.94
Adj-Rsquare	0.35				

表 3: 领先多因素模型（样本区间 2002.1-2006.12、2006.1-2010.12）

	样本区间：2002.1-2006.12			样本区间：2006.1-2010.12		
	常数项	PPI	储蓄/M2 同比增速	常数项	PPI	M1 同比增速
j		2	2		3	6
k		-2	-2		-2	-2
coefficient	0.84	-4.13	5.27	1.57	-8.37	8.01
t-test	0.48	-2.42	3.18	0.65	-3.41	3.24
Adj-Rsquare	0.23			0.35		

资料来源：Wind，长江证券研究部。

根据模型（3）对领先多因素模型的预测效果进行检验，我们发现系数 β_1 的 t 值不显著，仅为 0.10，整个模型的 Rsquare 也几乎为零。

图 1: 领先多因素模型样本外择时效果图 (2001.1-2011.10)



资料来源: Wind, 长江证券研究部。

依据我们一直以来的择时方法: 如果预测值为正则持有上证指数, 如果预测值为负则空仓, 我们也模拟了样本外的择时效果, 如上图所示。与数据检验的结果一致, 从图形上看择时的效果也较差, 尤其是在一些年份的后半年, 如 2006 年后半年, 2007 年后半年, 以及 2008 年后半年这几段时间择时效果较差。

2. 及时纳入新的领先变量对预测效果的改进

上一节我们看到领先多因素模型预测效果较差, 我们也提出, 从择时效果图 11 可以看到, 尤其是在一些年份的后半年, 如 2006 年后半年, 2007 年后半年, 以及 2008 年后半年这几段时间择时效果较差。

上述多因素模型对股市预判过程中, 我们坚持用一个模型预测未来 12 个月的收益率, 12 个月之后我们再更新模型, 纳入新出现的领先变量。我们清楚地知道, 宏观经济变量与股市的关系是不稳定的, 尤其是领先变量, 与股市的关系在近 10 年内发生了很大的改变, 我们的做法是总结过去一段时间的规律, 跟随规律, 并运用规律对未来的股市做出预判, 因此针对宏观经济变量与股市之间的这种不稳定关系, 我们需要用及时调整规律的心态来对待, 因此这一节, 我们探讨这个问题: 及时纳入新的领先变量是否会明显改进预测的效果, 是否会提高预测的准确性?

我们仍旧采取滚动外推的做法对样本外的股市进行预测, 但是调整了更新模型的频率, 将更新模型的频率从一年更新一次模型调整为一个月更新一次模型, 具体的做法是: 从 1996 年 1 月开始, 选取 5 年的数据 (1996 年 1 月至 2000 年 12 月, 共 60 个样本点) 检验宏观经济变量与股市的关系, 并找出领先的变量建立多因素模型, 并且对未来一个月, 即: 2001 年 1 月的股市收益率进行预判, 接下来, 加入这一个月新的宏观经济变量的数据, 仍旧保持用最新 5 年的数据建立新的领先多因素模型, 再对未来一个月的股市收益率进行预判, 以此方法逐步外推。采用这样的方法, 我们在 2000 年 1 月至 2010 年 10 月这 10 年内的每个月都会建立一个领先多因素模型, 并且根据模型预测了 2001 年 1 月至 2011 年 10 月每个月的股市收益率。仍旧使用模型 (3) 来判断多因素模型的样本外预测效果。

下面表 4 列举了样本区间分别选取 2005.10-2010.9 与 2006.1-2010.12 所建立的领先多因素模型，可以看到，虽然这两个结果的样本选择只有前后六个月的差异，但是建立的多因素模型是不一样的。

表 4：领先多因素模型（样本区间 2005.10-2010.9、2006.1-2010.12）

	样本区间：2005.10-2010.9			样本区间：2006.1-2010.12		
	常数项	PPI	短期贷款 增速	常数项	PPI	M1 同比增速
j		3	4		3	6
k		-2	-1		-2	-2
coefficient	-0.45	-7.34	9.05	1.57	-8.37	8.01
t-test	-0.15	-2.82	3.22	0.65	-3.41	3.24
Adj-Rsquare	0.36			0.35		

资料来源：Wind，长江证券研究部。

这样的情况并不是偶然，甚至在 2000 年到 2010 年这段时间内，我们发现有些变量与股市的关系发生了根本性的改变，比如在 1996 年至 2001 年这段时间，我们发现 PPI 是正向领先股市，而最近的五年 PPI 却是反向领先股市的。如表 5 所示。

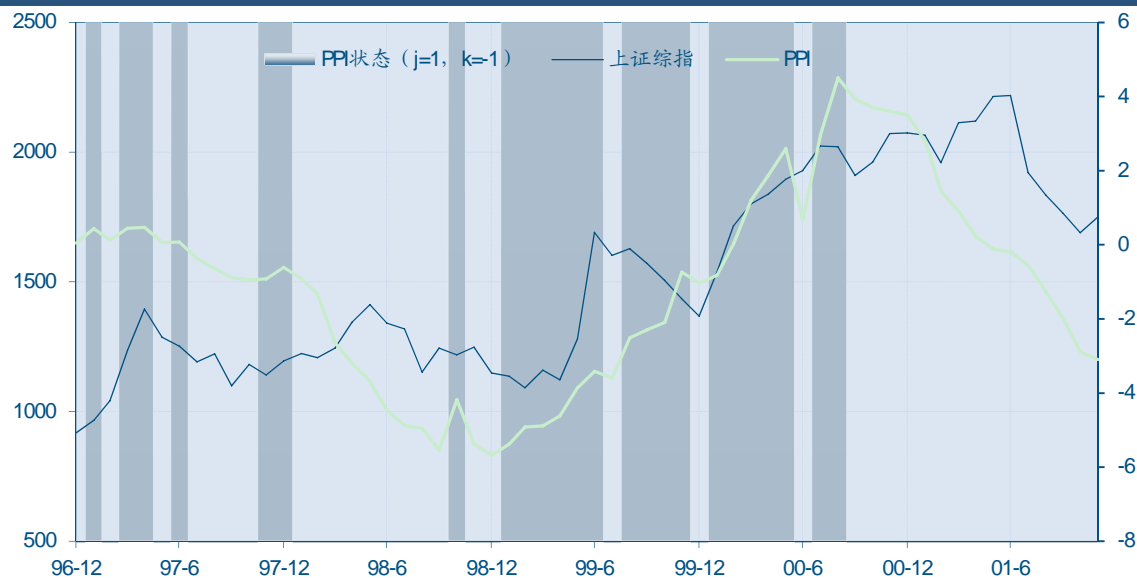
表 5：领先多因素模型（样本区间 1996.12-2001.11）

样本区间： 1996.12-2001.11	CPI	PPI	社会消费 品零售总 额同比	M0 同比增 速	短期贷款 同比
j	1	1	2	2	1
k	-2	-1	-2	-1	-2
coefficient	6.44	-5.07	6.21	-3.60	-5.36
t-test	3.65	-3.00	3.76	-2.04	-3.25
Adj-Rsquare	0.42				

资料来源：Wind，长江证券研究部。

下面是 1996 年 12 月到 2001 年 11 月 PPI 对上证综指的正向领先关系图。

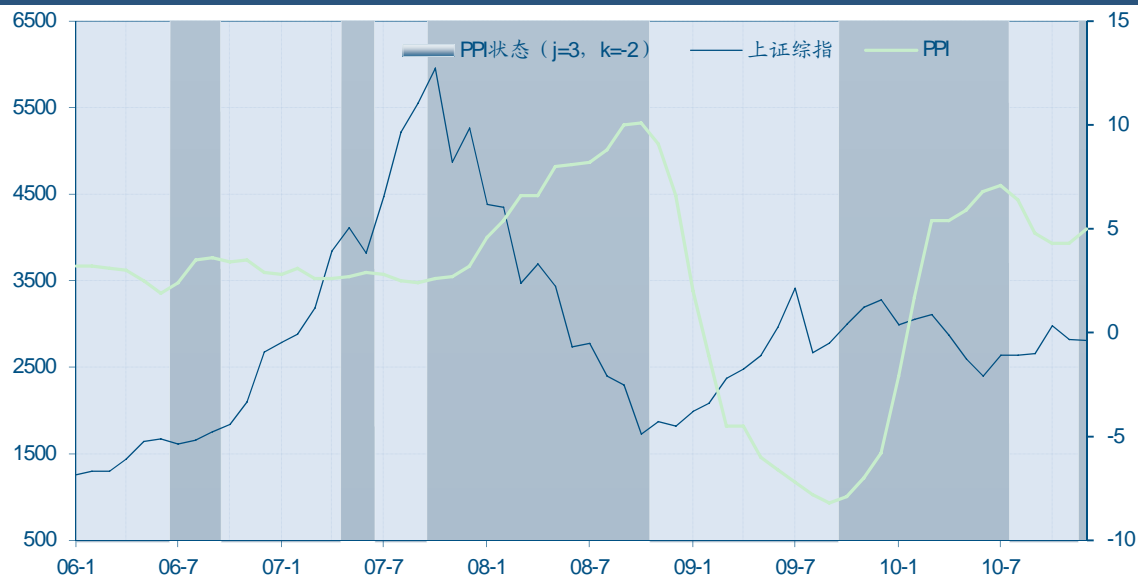
图 2: PPI 与上证综指的关系图 (1996.12-2001.11)



资料来源: Wind, 长江证券研究部。

下面是 2006 年 1 月到 2010 年 12 月 PPI 对上证综指的反向领先关系图。

图 3: PPI 与上证综指的关系图 (2006.1-2010.12)



资料来源: Wind, 长江证券研究部。

这种宏观经济变量与股市关系不稳定的现象说明, 提高更新模型的频率有助于及时有效地纳入新的领先变量, 进而尽快把握最新的宏观经济变量与股市的关系, 根据新的规律对未来的股市做出预判, 提高对股市预测的准确性。

下表 6 展示了及时纳入领先变量对预测效果的改进, 可以看出, 提高更新模型的频率对样本外的预测效果有明显改进, 及时纳入领先变量建立多因素模型样本外的预测与收益率与实际收益率有显著的正向关系。

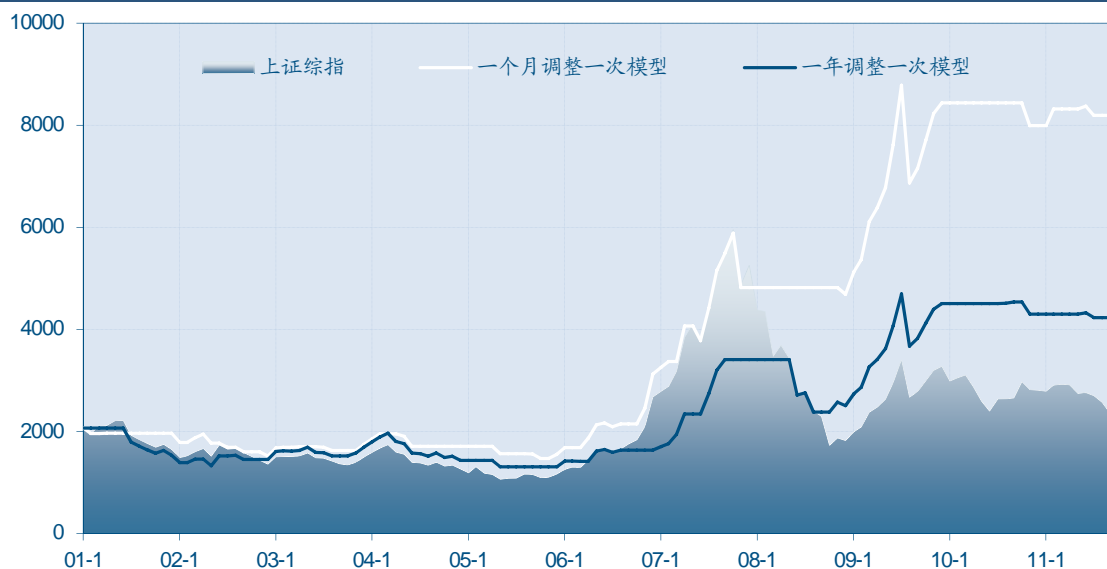
表 6: 样本外预测效果检验结果对比

	一年更新一次模型	一个月更新一次模型
coefficient	0.01	0.48
t-test	0.10	3.23
Rsquare	0.00	0.08

资料来源: Wind, 长江证券研究部。

及时纳入新的领先变量对多因素模型的样本外预测效果改进程度如如下图 11 所示, 一个月更新一次模型相比一年更新一次模型择时效果有明显的改善。

图 4: 领先多因素模型样本外择时效果图 (2001.1-2011.10)



资料来源: Wind, 长江证券研究部。

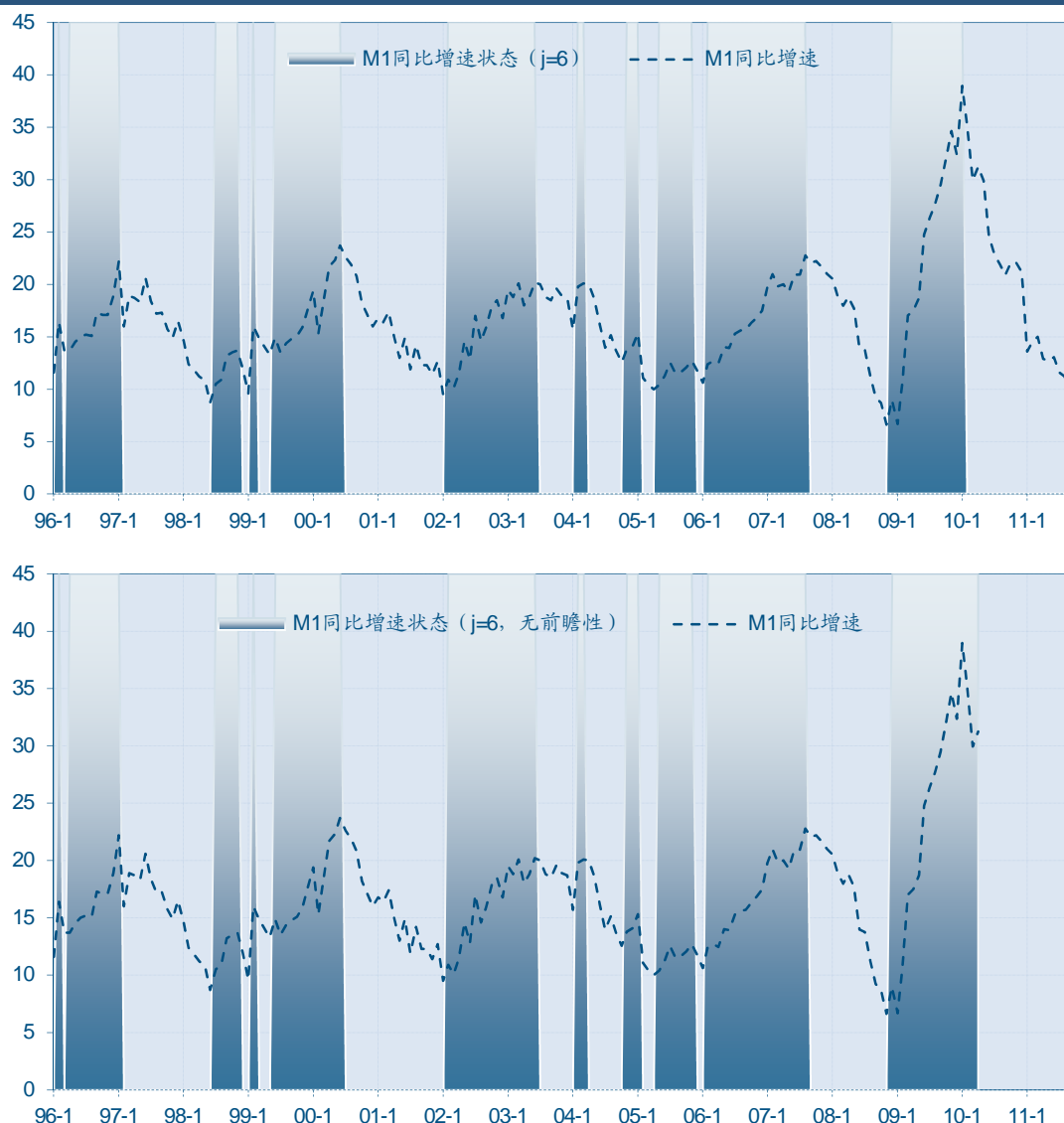
3. 前瞻性对预测效果的改进

在识别宏观经济变量的趋势状态时, 我们采用的是事后识别的办法, 具体的方法是先标出高低点, 再识别出上升下降趋势。由于在标记高低点的时候, 我们需要考察左侧及右侧的数据来确定当前的点是否为高点或低点, 而右侧的数据意味着是当前点未来的数据, 因此对于当前点是否是高点或者低点, 需要对数据有一些前瞻性才可以做出准确的判断, 也就是说当没有右侧的数据, 或者右侧的数据不足够时, 我们的方法对宏观经济变量的趋势状态的识别是可能与实际情况不相符合, 这个问题在实际的预判中是无法解决的, 除非我们加入自己对宏观经济变量未来趋势的主观预判。但是, 在研究过去的宏观经济变量与股市的关系时, 在对过去的宏观经济变量建立多因素模型, 并且研究其外推效果时, 我们可以来检验如果对宏观经济变量有一定的前瞻性, 也就是说在识别宏观经济变量的趋势状态时, 可以完全获得右侧的数据, 这样的方法得出的规律建立多因素模型, 在预测样本外股市收益时是否会有更

高的胜率？这是我们这一节要回答的问题。

图 5 中的两张图是识别 M1 增速的趋势时，有无前瞻性的对照，上面一张是识别趋势是具有前瞻性识别出来的结果，下面一张是仅考虑当前可以获得的数据，也就是左侧的数据识别出来的结果，可以看到最后一个高点出现在 2010 年 1 月，而在仅有 2010 年 4 月左侧的数据时无法识别出 2010 年 1 月这个高点之后已经开始的下降趋势。因此这一节我们主要讨论，增加对宏观经济变量的前瞻性是否可以改进模型的样本外预测效果。

图 5: M1 同比增速及其状态，有前瞻性与无前瞻性对比



资料来源：Wind，长江证券研究部。

我们仍旧采用一个月更新一次模型这种可以及时纳入新领先变量的方法外推预测，与前面唯一不同的地方在于：识别宏观经济变量的趋势时假设完全已知所有的数据，同样的，使用模型（3）来判断多因素模型的样本外预测效果。

下表 7、表 8 是对宏观经济变量的趋势没有前瞻性、对宏观经济变量的趋势具有前瞻性建立的多因素模型对比，样本区间是 2000 年 1 月至 2004 年 12 月、2006 年 1 月至 2010 年 12 月。

从这两个时间建立的模型来看,对宏观经济变量的趋势没有前瞻性、对宏观经济变量的趋势具有前瞻性建立的多因素模型包含的变量会有些区别,从解释能力的角度来看,对宏观经济变量的趋势具有前瞻性建立的多因素模型对股市解释能力较大。

表 7: 无前瞻性与有前瞻性识别趋势建立的多因素模型对比 (样本区间为 2000.1-2004.12)

	对宏观经济变量的趋势没有前瞻性			对宏观经济变量的趋势具有前瞻性		
样本区间: 2000.1-2004.12	常数项	工业增加值 值同比	M0 同比增 长率	常数项	工业增加值 值同比	发电量同 比
J		4	1		4	4
K		-2	-1		-2	-2
coefficient	-1.66	4.25	-2.70	-0.61	3.51	-3.53
t-test	-1.48	3.20	-2.05	-0.50	2.73	-2.78
Adj-Rsquare	0.16			0.21		

表 8: 无前瞻性与有前瞻性识别趋势建立的多因素模型对比 (样本区间为 2006.1-2010.12)

	对宏观经济变量的趋势没有前瞻性			对宏观经济变量的趋势具有前瞻性		
样本区间: 2006.1-2010.12	常数项	PPI	M1 同比增速	常数项	PPI	短期贷款 同比
J		3	6		3	4
K		-2	-2		-2	-1
coefficient	1.57	-8.37	8.01	-0.59	-7.18	8.88
t-test	0.65	-3.41	3.24	-0.19	-2.75	3.14
Adj-Rsquare	0.35			0.34		

资料来源: Wind, 长江证券研究部。

下表是根据模型 (3) 分别对宏观经济变量的趋势没有前瞻性、对宏观经济变量的趋势具有前瞻性建立的多因素模型外推效果的检验结果,从下面的结果可以看到,在识别趋势时,增加对宏观经济变量趋势的前瞻性会提高样本外预判的准确性。

表 9: 样本外预测效果对比

	一年更新一次模型		一个月更新一次模型	
识别状态时有无 前瞻性	无前瞻性	有前瞻性	无前瞻性	有前瞻性
coefficient	0.01	0.30	0.48	0.54
t-test	0.10	2.31	3.23	3.94

Rsquare

0.00

0.04

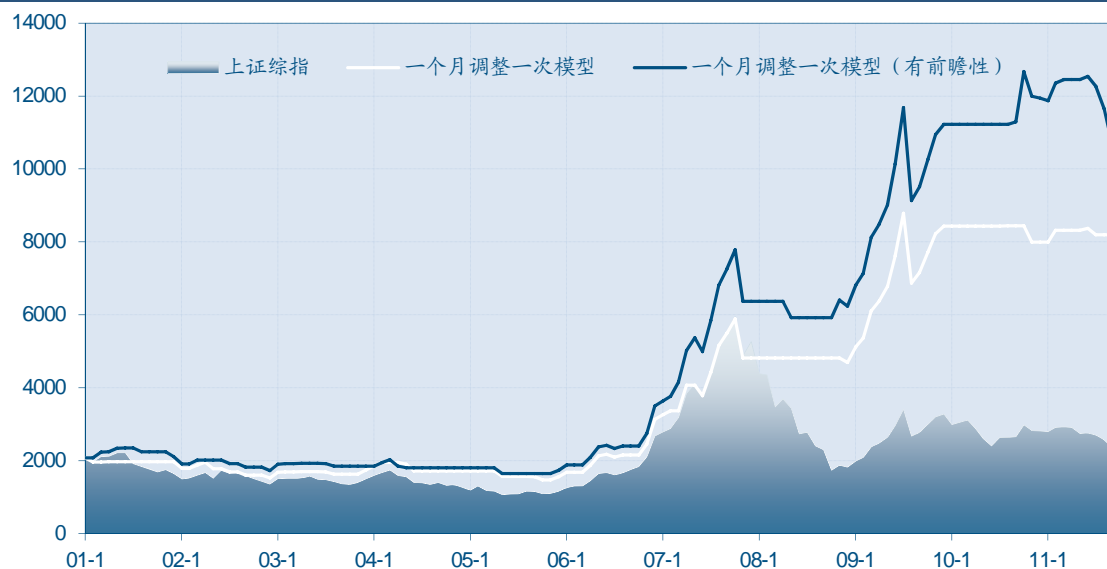
0.08

0.11

资料来源：Wind，长江证券研究部。

增加对宏观经济变量趋势的前瞻性建立的多因素模型的样本外预测效果改进程度如下图 6 所示，对宏观经济变量趋势的前瞻性可以明显改进模型的样本外择时效果。

图 6：多因素模型外推一年的择时效果



资料来源：Wind，长江证券研究部。

三、总结与思考

通过上面的研究，我们发现宏观经济变量对股市的预测效果是明显的，并且同时也发现有两个方面的因素可以提高宏观经济变量对股市的预测效果：

第一、及时纳入新的领先变量可以提高对股市的预测效果。我们使用 5 年的数据建立领先多因素模型，分别逐年、逐月外推对股市进行样本外预测，发现逐月外推相比逐年外推可以提高预测的准确性，样本外的预测方向正确的概率达到 58%。

第二、对宏观经济变量的趋势具有前瞻性可以提高对股市的预测效果。我们在对宏观经济变量的趋势具有前瞻性的情况下建立多因素模型，逐月外推对股市进行样本外预测，发现对宏观经济变量的趋势具有前瞻性的多因素模型可以提高预测的准确性，样本外的预测方向正确的概率达到 61%。

总的来说，宏观经济变量对股市的预测效果是明显的，我们发现宏观经济变量与股市的关系是不稳定的，因此需要及时纳入新的领先变量以把握二者的近期关系以提高预测胜率。另外，由于对宏观经济变量拐点的识别总会滞后于拐点发生的时刻，如果对宏观经济变量趋势具有一些前瞻性，那么就可以准确判断当前宏观经济变量的趋势，这样自然可以提高对股市的预测胜率。

这篇报告在《宏观经济变量与股市关系系列报告》之后对宏观经济变量与股市关系的一些思考与改进，希望我们可以在这个领域尽力完善自己的研究。

分析师介绍

范辛亭，中国科学技术大学博士，香港中文大学博士后，中山大学副教授，长江证券金融工程首席分析师。

武丹，华东师范大学数学系本科，复旦大学管理学院硕士，从事金融工程研究。

对本报告的评价请反馈至长江证券机构客户部

姓名	分工	电话	E-mail
伍朝晖	副主管	(8621) 68752398	13564079561 wuzh@cjsc.com.cn
甘 露	华东区总经理	(8621) 68751916	13701696936 ganlu@cjsc.com.cn
鞠 雷	华南区总经理	(8621) 68751863	13817196202 julei@cjsc.com.cn
程 杨	华北区总经理	(8621) 68753198	13564638080 chengyang1@cjsc.com.cn
李劲雪	上海私募总经理	(8621) 68751926	13818973382 lijx@cjsc.com.cn
张 晖	深圳私募总经理	(0755) 82766999	13502836130 zhanghui1@cjsc.com.cn

投资评级说明

行业评级	报告发布日后的 12 个月内行业股票指数的涨跌幅度相对同期沪深 300 指数的涨跌幅为基准，投资建议的评级标准为： 看好： 相对表现优于市场 中性： 相对表现与市场持平 看淡： 相对表现弱于市场
公司评级	报告发布日后的 12 个月内公司的涨跌幅度相对同期沪深 300 指数的涨跌幅为基准，投资建议的评级标准为： 推荐： 相对大盘涨幅大于 10% 谨慎推荐： 相对大盘涨幅在 5%~10%之间 中性： 相对大盘涨幅在-5%~5%之间 减持： 相对大盘涨幅小于-5% 无投资评级： 由于我们无法获取必要的资料，或者公司面临无法预见结果的重大不确定性事件，或者其他原因，致使我们无法给出明确的投资评级。

重要声明

长江证券股份有限公司具有证券投资咨询业务资格，经营证券业务许可证编号：Z24935000。

长江证券系列报告的信息均来源于公开资料，本公司对这些信息的准确性和完整性不作任何保证，也不保证所包含的信息和建议不会发生任何变更。本公司已力求报告内容的客观、公正，但文中的观点、结论和建议仅供参考，不代表对证券价格涨跌或市场走势的确定性判断。报告中的信息或意见并不构成所述证券的买卖出价或征价，投资者据此做出的任何投资决策与本公司和作者无关。

本公司及作者在自身所知情的范围内，与本报告中所评价或推荐的证券没有利害关系。本公司及其所属关联机构可能会持有报告中提到的公司所发行的证券头寸并进行交易，也可能为这些公司提供或者争取提供投资银行、财务顾问或者金融产品等相关服务。

本报告版权仅为本公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布。如引用、刊发，需注明出处为长江证券研究部，且不得对本报告进行有悖原意的引用、删节和修改。