

Federal Reserve Bank of New York  
Staff Reports

Reexamining the Consumption-Wealth Relationship:  
The Role of Model Uncertainty

Gary Koop  
Simon M. Potter  
Rodney W. Strachan

Staff Report no. 202  
March 2005

This paper presents preliminary findings and is being distributed to economists and other interested readers solely to stimulate discussion and elicit comments. The views expressed in the paper are those of the authors and are not necessarily reflective of views at the Federal Reserve Bank of New York or the Federal Reserve System. Any errors or omissions are the responsibility of the authors.

## **Reexamining the Consumption-Wealth Relationship: The Role of Model Uncertainty**

Gary Koop, Simon M. Potter, and Rodney W. Strachan

*Federal Reserve Bank of New York Staff Reports*, no. 202

March 2005

JEL classification: C11, E21

### **Abstract**

In their influential work on the consumption-wealth relationship, Lettau and Ludvigson found that while consumption responds to permanent changes in wealth in the expected manner, most changes in wealth are transitory with no effect on consumption. We investigate the robustness of these results to model uncertainty using Bayesian model averaging. We find that there is model uncertainty with regard to the number of cointegrating vectors, the form of deterministic components, lag length, and whether the cointegrating residuals affect consumption and income directly. Whether this uncertainty has important implications depends on the researcher's attitude toward this economic theory used by Lettau and Ludvigson. If we work with their exact model, our findings are very similar. However, if we work with a broader set of models, we find that the exact magnitude of the role of permanent shocks is difficult to estimate precisely. Thus, although some support exists for the view that the role of shocks is small, we cannot rule out the possibility that they have a substantive effect on consumption.

Key words: wealth effect, vector error correction model, Bayesian model averaging, cointegration, variance decomposition

---

Koop: Department of Economics, University of Leicester (e-mail: Gary.Koop@leicester.ac.uk).  
Potter: Federal Reserve Bank of New York (e-mail: Simon.Potter@ny.frb.org). Strachan:  
Department of Economics, University of Leicester (e-mail: rws7@le.ac.uk). The authors thank  
Sydney Ludvigson and seminar participants at the University of Birmingham for helpful  
comments. The views expressed here are those of the authors, and do not necessarily reflect  
the position of the Federal Reserve Bank of New York or the Federal Reserve System.

# 1 Introduction

Textbook wisdom suggests that the size of the wealth effect (i.e. the change in consumption induced by a \$1 increase in wealth) should be approximately 5 cents.<sup>1</sup> However, recent events which caused large changes in wealth without commensurately large changes in consumption (e.g. the stock market crash of the October 1987 and the rise and fall of stock prices associated with the “New Economy” of the 1990s) have raised the issue of whether the wealth effect really is this large. This uncertainty over the magnitude of the wealth effect has stimulated a large amount of recent research. An important paper, Lettau and Ludvigson (2004) (LL hereafter), presents empirical evidence which attempts to reconcile the apparent conflict between textbook wisdom and recent experience. Using cointegration techniques, LL estimate permanent and transitory components of the fluctuations in wealth. They find that consumption does respond to permanent changes in wealth in the textbook fashion, but does not respond to transitory changes in wealth. However, most of the fluctuations in wealth are transitory. In essence, the “Dotcom bubble” had little effect on consumption since its effects on wealth were correctly anticipated to be transitory.

Like many empirical conclusions in economics, LL’s are based on the estimated properties of a single econometric model, selected after a long and careful specification search. For many years, economists and statisticians have been worried about such a strategy which treats the final selected model as though it were the true and ignores the evidence from other models. For instance, Leamer (1978) presents a persuasive argument for basing empirical conclusions on all possible models under consideration, instead of selecting one model through a data-driven specification search. In the statistical literature, Draper (1995) and Hodges (1987) make similar calls for model uncertainty to be reflected in policy analysis. Macroeconomic contributions such as Cogley and Sargent (2003), Fernandez, Ley and Steel (2001) and Sala-i-Martin, Doppelhoffer and Miller (2004) have convinced many of the importance of model uncertainty for empirical practice. Most such papers adopt a Bayesian perspective since it treats models as random variables and, thus, averaging across models can be done in a logical and coherent manner.<sup>2</sup> The basic idea of Bayesian model averaging can be explained quite simply. Suppose the

---

<sup>1</sup>Lettau and Ludvigson (2004) present results for a wide sample of macroeconomics textbooks indicating support for values in this region.

<sup>2</sup>Papers which investigate the use of non-Bayesian model averaging techniques in macroeconomic applications include Garratt, Lee, Pesaran and Shin (2003), Min and Zellner (1993) and Sala-i-Martin (1997). An example of a non-Bayesian theoretical discussion of model averaging is Hjort and Claeskens (2003).

researcher is entertaining  $R$  possible models, denoted by  $M_1, \dots, M_R$ , to learn about an unknown feature of the economy,  $\phi$  (e.g. a variance decomposition or an impulse response to a particular shock). If we treat  $\phi$  and  $M_r$  as random variables, the rules of conditional expectation imply that:

$$E(\phi|Data) = \sum_{r=1}^R p(M_r|Data) E(\phi|Data, M_r). \quad (1.1)$$

Thus, overall point estimates,  $E(\phi|Data)$ , should be an average of point estimates in individual models,  $E(\phi|Data, M_r)$ .<sup>3</sup> The weights in the average are the posterior model probabilities,  $p(M_r|Data)$ . There are many other justifications for using model averaging, ranging from a desire to avoid the well-known pre-test problem [e.g. Poirier (1995), pages 519-523] to formal justifications of such an approach in decision theoretic contexts [e.g. Min and Zellner (1993) or Raftery, Madigan and Hoeting (1997)].

Given the importance for macroeconomic policy of LL's results and the large and growing literature using a similar specification [see Julliard (2005)], it is crucial to investigate how robust their results are with respect to plausible changes in model assumptions and the Bayesian framework is a logical one in which to approach this problem. This is what we do in this paper. Using Bayesian reference prior methods for cointegrated models developed in a series of papers [Strachan (2003), Strachan and Inder (2004) and Strachan and van Dijk (2003)], we carry out a Bayesian model averaging exercise with the data used in LL where different models are defined depending on the number of cointegrating vectors, form of deterministic trends, lag length and whether the cointegrating residuals affect consumption and income directly. We find that many different models receive appreciable support from the data and, thus, substantial model uncertainty exists. As a result of this, we find that the exact magnitude of the role of permanent shocks is hard to estimate precisely. Thus, although some support exists for the view that their role is small, we cannot rule out the possibility that they have a substantive role to play. In particular, the posterior distributions of key variance decompositions are found to be multimodal and relatively non-informative. However, when we consider only the single model used by LL with the cointegrating residual only directly allowed to effect wealth (and not consumption or income), then our results are similar to theirs.

---

<sup>3</sup>The same argument holds for other features of the posterior distribution (e.g. predictives or measures of dispersion such as Highest Posterior Density Intervals).

## 2 Cointegration and the Consumption-Wealth Relationship

LL, using derivations from an earlier paper [Lettau and Ludvigson (2001)], base their empirical analysis of the cointegration-wealth relationship on an accounting relationship based on the budget constraint linking tomorrow’s wealth to today’s wealth, consumption and the rate of return on wealth. Because wealth contains a human capital component it is not directly observable. Accordingly, LL log linearize the budget constraint relationship, assume consumption growth and the returns on human capital and asset wealth are stationary and relate the nonstationary component of human capital to labor income. They end up with a model where log consumption,  $c_t$ , log asset wealth,  $a_t$ , and log labor income,  $y_t$ , should be cointegrated. LL focus on the case of a single cointegrating relationship they call “cay”,  $c_t - \alpha_a a_t - \alpha_y y_t$ . That is, the cointegrating vector should be  $(1, -\alpha_a, -\alpha_y)'$  and there should be no deterministic trend in the cointegrating residual (although there may be an intercept). As LL show, this implies that there are two permanent shocks and one transitory shock driving the joint fluctuations in consumption, labor income and wealth. Their accounting argument does not rule out the case of two cointegrating relationships, with the implication of one permanent and two transitory shocks and once again no deterministic trends in the cointegrating residuals. Note that in this latter case “cay” is still stationary. In theory, the budget constraint should also imply  $\alpha_a + \alpha_y = 1$  but, given that only a fraction of total consumption based on nondurables and services is observable, we should get  $0 \leq \alpha_a + \alpha_y \leq 1$ .

An important methodological issue is how to include the theoretical information outlined in the previous paragraph in the statistical model. From a Bayesian point of view, this can be thought of as prior information, either about the model space (e.g. the model with one cointegrating vector involving no deterministic trend in the cointegrating residual is more plausible *a priori*) or about the parameter space (e.g. the region  $0 \leq \alpha_a + \alpha_y \leq 1$  is more plausible *a priori*). In this paper, we use a Bayesian approach to re-examine the conclusions of LL about the consumption-wealth relationship. In this re-examination, we view Bayesian methods as a tool that allows us to investigate the relative roles of economic theory, the data and the model selection procedure in empirical conclusions. In particular, in the re-examination we conduct, there are substantial economic theory reasons to adopt a number (but not all) of the model specification choices used by LL. What we wish to do is separate these *a priori* beliefs from the information in the observed data and their data-driven model specification choices. This is not because we believe that strong theory-driven prior beliefs in favor of certain models are necessarily

bad. On the contrary we believe that such prior information is a necessary but not sufficient condition for obtaining robust empirical conclusions. However, it is important to understand the relative contributions of *a priori* beliefs, data-driven modeling choices and the data itself.

These distinctions can be made clear by describing our class of econometric models. This class is a standard one as outlined in, e.g., Johansen (1995). If we let  $x_t$  be an  $n \times 1$  vector (i.e. in our case  $n = 3$  and  $x_t = (c_t, a_t, y_t)'$ ), then a Vector Error Correction model (VECM) can be written as:

$$\Delta x_t = \gamma \alpha' x_{t-1} + d_t \mu + \Gamma(L) \Delta x_{t-1} + \varepsilon_t, \quad (2.1)$$

where  $\gamma$  and  $\alpha$  are  $n \times r$  matrices with  $0 \leq r \leq n$  being the number of cointegrating relationships.<sup>4</sup>  $\Gamma(L)$  is a matrix polynomial of degree  $p$  in the lag operator and  $d_t$  is the deterministic term (to be defined shortly). The framework described in (2.1) defines a set of models which differ in the number of cointegrating vectors ( $r$ ), lag length ( $p$ ) and the specification of deterministic terms. With regards to the latter, a deterministic trend in the cointegrating residuals ( $\alpha' x_{t-1}$ ) has very different implications than a deterministic trend in the levels of the series. Accordingly, following Johansen (1995, Section 5.7), we decompose  $\mu$  into these two different parts as:

$$\mu = \gamma \mu_1 + \gamma_\perp \mu_2$$

where  $\mu_1 = (\gamma' \gamma)^{-1} \gamma' \mu$  and  $\mu_2 = (\gamma_\perp' \gamma_\perp)^{-1} \gamma_\perp' \mu$  (and  $\gamma_\perp$  is orthogonal to  $\gamma$ ). With this transformation,  $\mu_1$  reflects deterministic terms in the cointegrating residual while  $\mu_2$  reflects those in  $\Delta x_t$ . Within this framework we consider possibilities for the deterministic term from the set  $d_t = (1, t)$ ,  $d_t = (1)$  and  $d_t = (0)$  (i.e. no deterministic terms). Note that we do not rule out including a linear trend in the cointegrating residual as this may be reasonable in a finite sample if the consumption-wealth ratio is changing over time and, indeed, support for such a specification is found by Lee (2002). Accordingly we consider the five possibilities defined in the following table.<sup>5</sup>

---

<sup>4</sup>If  $r = n$  then  $\alpha = I_n$  and all the series do not have unit roots (i.e. this usually means they are stationary to begin with).

<sup>5</sup>See Johansen (1995), pages 81-84 for a motivation for these particular five choices (which involves eliminating some potentially observationally equivalent combinations of determinist terms).

Code	In Cointegrating Residual	In $\Delta x_t$
$d = 1$	$d_t = (1, t)$	$d_t = (1, t)$
$d = 2$	$d_t = (1, t)$	$d_t = (1)$
$d = 3$	$d_t = (1)$	$d_t = (1)$
$d = 4$	$d_t = (1)$	$d_t = (0)$
$d = 5$	$d_t = (0)$	$d_t = (0)$

Limiting the number of variables affected by the cointegrating residuals (in econometric language these are referred to as weak exogeneity restrictions) is another issue that is commonly considered in this literature. In some of their results, LL impose weak exogeneity of  $c_t$  and  $y_t$  in the cointegrating relationship. That is, they set the coefficients in  $\gamma$  corresponding to the equations for  $c_t$  and  $y_t$  to be zero. They argue, following Gonzalo and Ng (2001), that this will allow for more stable estimates of the permanent-transitory decomposition. The instability is produced by the sensitivity of  $\gamma_\perp$  to estimation error in  $\gamma$ . We use an alternative technique, due to Centoni and Cubadda (2003) that does not require the construction of  $\gamma_\perp$  to measure the share of permanent shocks in the variance decomposition.<sup>6</sup> This form of weak exogeneity has the important policy implication that transitory shocks have no immediate effect on  $c_t$  and  $y_t$  and, at longer horizons, the effect of transitory shocks enters through the asset channel's affect on consumption and labor income. For example, if this form of weak exogeneity is imposed and the lags of wealth growth in the consumption and labor income equations are all close to zero, then nearly all of the adjustment to the transitory shock will be in wealth.

We have now defined a set of models (i.e. a model space) which vary in their number of cointegrating vectors ( $r$ ), lag length ( $p$ ), deterministic terms ( $d$ ) and whether or not weak exogeneity is imposed. Economic theory has strong implications for some aspects of model choice, but not others. In particular, it says that we should have  $r = 1$  or  $2$  and the cointegrating residual should not have a trend in it (i.e.  $d > 2$ ). However, economic theory says nothing about the deterministic term in  $\Delta x_t$ , nor about lag length nor about weak exogeneity. One can imagine three different reactions to this state of affairs. Firstly, one could simply argue that theory should be ignored and we should let the data speak. This argues for the consideration of all possible models and a non-informative prior over model space. Secondly, one could argue that theory should be imposed dogmatically and, thus, only models with  $r = 1$  or  $2$  and  $d > 2$  should be considered. This argues for a non-informative prior over the set of models consistent with theory. Thirdly, one could argue that all models should be considered, but

<sup>6</sup>Based on simulation comparisons between the two approaches we suspect that the two ostensibly different methods produce identical results for the variance decomposition presented later.

that more weight should be attached to models which are consistent with theory. This argues for an informative (but not dogmatic) prior over model space. In our empirical work, we discuss all three of these strategies.

Our econometric methods, which are largely the same as those developed in Strachan (2003), Strachan and Inder (2004) and Strachan and van Dijk (2003), are described in the Appendix. Here we sketch the basic ideas. The key elements we require are a posterior for the parameters and a posterior model probability for each model. In terms of (1.1), to obtain posterior properties of a function of interest (e.g. a variance decomposition), we need  $p(\phi|Data)$  which depends on  $p(M_r|Data)$  and  $p(\phi|Data, M_r)$ . Properties of the latter two densities can be obtained from the posterior simulators for each model described in the Appendix.  $p(M_r|Data)$  and  $p(\phi|Data, M_r)$  depend on the likelihood function (defined by equation 2.1, assuming errors are Normally distributed), a prior over the parameter space and a prior over model space. The prior over model space will be discussed in the next section. With regards to the prior over parameter space we adopt three different approaches. First, we use a standard noninformative prior but use the Schwarz criterion (BIC) to approximate  $p(M_r|Data)$ .<sup>7</sup> Secondly, we use a standard noninformative prior and the fractional Bayes factor approach of O’Hagan (1995). Thirdly, we use an informative shrinkage prior similar in spirit to the commonly-used shrinkage prior of Doan, Litterman and Sims (1984). However, in order to be as objective as possible, we treat the shrinkage parameter as an unknown parameter updated in a data-based fashion.<sup>8</sup> For the prior over the cointegration space, we choose the reference prior described in Strachan and Inder (2004).<sup>9</sup> This prior is Uniform over cointegration space and, thus, noninformative in the sense described in Strachan and Inder (2004). In addition, we present results averaged over the three priors. Such results can be motivated by noting that a Bayesian model involves both a likelihood and a prior. Interpreted in this way, our empirical work involves a huge set of models defined not only by likelihood assumptions (e.g. number of cointegrating vectors), but also by prior assumptions and it makes sense to present a grand average over all such models.<sup>10</sup> Thus, we are using a

---

<sup>7</sup>The ratio of BIC’s for two models is approximately equal to the log of the Bayes factor comparing the two models. This relationship can be used to approximate  $p(M_r|Data)$ .

<sup>8</sup>In Bayesian jargon, we use a hierarchical prior for the shrinkage parameter.

<sup>9</sup>A reference prior is one which is can be used automatically, not requiring subjective prior input from the researcher. For the purposes of model comparison, care must be taken in designing a reference prior to ensure that all models receive equal treatment. In the Bayesian cointegration literature, there has been a long discussion as to what makes a good reference prior. This is not the place to summarize the issues raised in this debate. Suffice it to note that the approach used in this paper surmounts many problems of earlier approaches and shares many similarities with other popular reference priors such as those developed in Kleibergen and Paap (2002) and Villani (2004).

<sup>10</sup>We use a simple average with equal weights for the BIC, Fractional Bayes factor and shrinkage prior approaches. Since the BIC approach is an approximation ignoring the prior, using marginal likelihoods as weights would not be possible.



wide variety of objective and reference Bayesian approaches to model comparison. We argue such an approach is crucial in order to investigate the prior robustness of any empirical results. Precise details are provided in the Appendix.

### 3 Empirical Results

Our empirical results are based on the same data as used in LL and the reader is referred there for precise details. Briefly, our data run from 1951Q4 through 2003Q1<sup>11</sup> and contains data on  $c_t$  which is the log of real per capita expenditures on nondurables and services (excluding shoes and clothing);  $a_t$  which is the log of a measure of real per capita household net worth (including all financial and household wealth as well as consumer durables); and  $y_t$  which is the log of after-tax labor income.

#### 3.1 Model Comparison Results

In this section, we present results relating to the posterior model probabilities,  $p(M_r|Data)$ , for our 80 different models (or  $3 \times 80$  if our three different classes of prior are interpreted as defining different model classes). As described in the previous section, models are defined by the number of cointegrating vectors ( $r = 0, 1, 2, 3$ ), the number of lags ( $p = 0, 1, 2, 3$ ) and the treatment of deterministic terms as outlined in Table 1 ( $d = 1, \dots, 5$ ).<sup>12</sup> For comparison, remember that LL chose  $r = 1$ ,  $d = 3$  and  $p = 1$ . As an aside, it is worth noting that, in their Table B.1, LL select  $p = 1$  using BIC (and their variance decompositions are based on  $p = 1$  and  $r = 1$ ), but they mention that AIC selects  $p = 0$ . Most of their statistical tests indicate  $r = 1$ . However, for  $p = 1$ , the Johansen L-max and trace tests in their Table B.1 both select  $r = 0$ . Hence, even in LL’s non-Bayesian procedure there does seem to be model uncertainty over both the number of cointegrating vectors and lag length.<sup>13</sup>

Given the large number of models and three different approaches to model comparison (i.e. based on BIC, fractional Bayes factors and the shrinkage prior), we do not present results for every model. Instead, the various sub-panels of Table 2 present results relating to the number of cointegrating vectors (integrating over

<sup>11</sup>Note that we also worked with the 1951Q4 through 1995Q4 sub-sample considered by LL. These sub-sample results are qualitatively similar to those obtained using the full sample and, for the sake of brevity, are not discussed here.

<sup>12</sup>Note that, in some cases, two different models are observationally equivalent (e.g. if  $r = 0$ , the treatment of deterministic terms in the cointegrating residual is irrelevant since a cointegrating residual does not exist). However, we leave observationally equivalent models in to make sure each value for  $r$  receives equal weight, *a priori*.

<sup>13</sup>LL use a different notation for lags so that their “2 lags” refers to 2 lags in the unrestricted VAR. This corresponds for our  $p = 1$  for lagged differences in the VECM. LL do not provide a detailed presentation of results relating to choice of deterministic terms.

deterministic terms and lag length), lag length (integrating over  $r$  and  $d$ ) and deterministic trends (integrating over  $r$  and  $p$ ) for our three different priors. In this table, we do not attach any extra weight to models consistent with economic theory, but simply allocate equal prior weight to each model.

With regards to the shrinkage prior, as described in the Appendix, this requires the choice of a shrinkage parameter which we call  $\eta$ . Exact details are given in the Appendix, but note here that we use a relatively diffuse hierarchical prior for this parameter (i.e. we are averaging over different values of  $\eta$  where the weights in the averaging are data-based).

Table 2: Results using Uniform prior over Space of All Models				
	BIC	Fractional Bayes Fact.	Shrinkage Prior	Average Over All Approaches
Results for Cointegrating Rank: $p(r Data)$				
$r = 0$	0.119	0.012	0.015	0.049
$r = 1$	0.878	0.273	0.283	0.478
$r = 2$	0.003	0.407	0.702	0.371
$r = 3$	0.000	0.309	0.000	0.103
Results for Lag Length: $p(p Data)$				
$p = 0$	0.996	0.010	0.000	0.335
$p = 1$	0.004	0.990	0.009	0.334
$p = 2$	0.000	0.001	0.043	0.014
$p = 3$	0.000	0.000	0.948	0.316
Results for Deterministic Trends: $p(d Data)$				
$d = 1$	0.000	0.017	0.000	0.006
$d = 2$	0.139	0.631	0.002	0.257
$d = 3$	0.861	0.091	0.444	0.465
$d = 4$	0.000	0.260	0.136	0.132
$d = 5$	0.000	0.002	0.418	0.140

A finding that leaps out is that there is a great deal of model uncertainty. For any given prior, there typically exists uncertainty over the number of cointegrating vectors and deterministic trends. Comparing across different priors we find even more uncertainty. Except for BIC, there is little evidence in favor of the single model used in LL. At various points, LL consider different values for lag length. Even if we integrate over lag length, the support for their model is not that strong with the shrinkage prior and fractional Bayes factor. That is, we find  $p(r = 1, d = 3|Data) = 0.740, 0.073$  and  $0.133$  using the BIC, fractional Bayes factor and shrinkage prior approaches, respectively. Only using BIC do we find LL's model to be the most probable one. The fractional Bayes factor and Shrinkage Prior approaches puts most probability on  $r = 2, d = 2, p = 1$  and  $r = 2, d = 5, p = 3$ ,

respectively.

With regards to cointegrating rank, BIC offers fairly strong support for the LL choice of  $r = 1$ . However, even BIC indicates an 11.9% chance that cointegration does not occur. Results using our other two priors are even more interesting, with under 30% chance of a single cointegrating vector, but appreciable weight allocated to two cointegrating vectors. The fractional Bayes factor approach even indicates appreciable support for  $r = 3$ . This latter finding, along with the appreciable support for  $d = 2$ , indicates support for the series being stationary with linear deterministic trends in their levels.

With regards to the issue of deterministic trends, there is also a high degree of uncertainty. Certainly, all of our approaches indicate substantial support for  $d = 3$  or 4 (the choice of LL and a closely related choice), but the BIC and fractional Bayes factor approaches attach substantial weight to the (economically non-intuitive) cases where there is a time trend in the cointegrating residual.

As we will discuss in more detail below, LL's finding of one cointegrating vector is crucial to their transitory-permanent decomposition which underlies many of their key empirical results. If all series have unit roots, but cointegration does not occur (i.e.  $r = 0$ ), then all shocks are permanent and their finding that transitory shocks to wealth are very important cannot be recovered. In contrast, if  $r = 3$  then all shocks are, by definition, temporary. Furthermore, we expect that models with  $r = 1$  should yield results similar to those found by LL, but there is no reason to think  $r = 2$  will necessary do so. When we do our Bayesian Model Averaging (BMA) exercise in the following section, we are averaging over all these choices so a wide variety of findings are possible.

The previous results are all based on a non-informative prior over model space. That is, each of our 80 models received an *a priori* weight of  $\frac{1}{80}$ . Given that economic theory suggests that cointegration is present ( $r = 1$  or 2) and  $d > 2$  (i.e. theory suggest that there is no linear trend in the consumption-wealth ratio as, in the long run, such a trend would imply values of this ratio outside the interval  $[0, 1]$ ), the researcher may wish to attach more prior weight to these models. For the sake of brevity we do not present results for this case, but stress that such an approach is possible. For instance, if the researcher thought that models consistent with economic theory should receive twice the weight of other models, then she should attach prior weight of  $\frac{2}{104}$  to the 24 models consistent with economic theory and  $\frac{1}{104}$  to the 56 models which are not. Results in Table 2 could be adjusted appropriately.

A third strategy a researcher might take is to impose economic theory on the model and work only with the

set of models that are consistent with it. If we do this, by attaching equal prior weight to all models with  $r = 1$  or  $2$  and  $d > 2$  (but are agnostic over lag length) and omit all other models, we obtain the results in Table 3. These results are more consistent with the findings of LL. Note, however, that with the exception of the BIC, there still is substantial probability attached to two cointegrating vectors and deterministic terms different from the ones used by LL. We will see whether these differences have substantial economic implications in the next section on variance decompositions.

Table 3: Results using Uniform prior over Models Consistent with Economic Theory				
	BIC	Fractional Bayes Fact.	Shrinkage Prior	Average Over All Approaches
Results for Cointegrating Rank: $p(r Data)$				
$r = 1$	0.997	0.341	0.720	0.686
$r = 2$	0.003	0.659	0.280	0.314
Results for Lag Length: $p(p Data)$				
$p = 0$	1.000	0.002	0.000	0.334
$p = 1$	0.000	0.998	0.009	0.336
$p = 2$	0.000	0.000	0.042	0.014
$p = 3$	0.000	0.000	0.949	0.316
Results for Deterministic Trends: $p(d Data)$				
$d = 3$	1.000	0.346	0.117	0.488
$d = 4$	0.000	0.647	0.263	0.297
$d = 5$	0.000	0.007	0.620	0.209

When doing variance decompositions, researchers sometimes impose weak exogeneity restrictions to improve estimation accuracy. In the present application, the weak exogeneity assumption of most interest (which is imposed by LL in some of their results) is that the coefficients in  $\gamma$  corresponding to the equations for  $c_t$  and  $y_t$  to be zero. This hypothesis only makes sense when cointegration occurs and  $r = 1$ . When  $r = 0$  these coefficients do not enter the model and when  $r = 2$  then cointegration theory implies that at most 1 variable can be weakly exogenous. Accordingly, we note that the probability of this weak exogeneity restriction, conditional on  $r = 1$ , is 0.398, 0.641 and 1.000 using the BIC, fractional Bayes factor and shrinkage prior approaches, respectively. Thus, there is some uncertainty about whether this is a reasonable restriction to impose.

Another hypothesis of interest is whether  $0 \leq \alpha_a + \alpha_y \leq 1$ . This hypothesis only makes sense when cointegration occurs with  $r = 1$ . As an example of our findings for this relationship, we find  $p(0 \leq \alpha_a + \alpha_y \leq 1|Data, r = 1)$  to be 0.614 using the shrinkage prior approach. Thus, there is substantial uncertainty over whether this implication of economic theory holds.

## 3.2 Variance Decompositions

We have established above that there is a great deal of uncertainty about which model is appropriate for this data set. Of course, if all of the plausible models have similar implications for quantities of interest to economists, then the fact that model uncertainty is substantial would be unimportant. Accordingly, in this section we investigate the consequences of model uncertainty for functions of interest to the economic policymakers. For the sake of brevity, we focus on variance decompositions. LL present a wider battery of results, but their variance decompositions are at the heart of their story. Briefly, cointegration restrictions allow us to decompose  $x_t$  into permanent and transitory components. We can then measure the role each of these plays in the model. A common way of doing this is to calculate the fraction of the total variance of the forecast error at horizon  $h$  which is attributable to the permanent and transitory components. LL provide more discussion. Note that, with three variables, we have three innovations driving the model (e.g. with  $r = 1$ , two of the innovations are permanent and one transitory). Following LL, we combine permanent shocks together (or transitory shocks together if  $r = 2$  ).

The key finding of LL is that transitory shocks dominate changes in wealth. Hence, although we will mention a range of findings, our main focus will be on this relationship. Furthermore, since our variance decompositions are measured as fractions of the total forecast error variance, the results for permanent and transitory shocks will sum to one. Accordingly, we only present variance decompositions for the permanent component. Finally, empirical results are qualitatively similar at all forecast horizons. Hence, we only present results for  $h = \infty$ .

Table 4 summarizes the posterior properties of the variance decomposition. We present the posterior mean and median (two commonly-used point estimates) as well as two measures of the dispersion of the posterior. These are a 50% Highest Posterior Density Interval (HPDI) which is defined as the shortest interval containing 50% of the posterior probability and an interquartile range (IQR): the 25th and 75th percentiles of the posterior.<sup>14</sup> At first glance, the reader may find the numbers in this table confusing. After all, for well-behaved distributions like the Normal, the mean and the median are the same and the 50% HPDI and the interquartile range should be as well. Clearly they are not. In fact, the HPDIs are often discontinuous, made up of two or more disjoint intervals. This property is due to the fact that BMA involves averaging different distributions

---

<sup>14</sup>We calculate these by creating a histogram defined over a grid and then using its properties to derive HPDIs and IQRs.

– the result can be a highly multimodal distribution. Note also that the model with  $r = 0$  implies all shocks are permanent. Hence, the permanent error’s share of the forecast error is one by definition. This will cause a spike in the posterior distribution at 1. If models with  $r = 0$  receive appreciable weight, this point can appear in HPDIs. So, for instance, the 50% HPDI interval for wealth using the BIC prior contains two intervals: one close to zero and one close to one. If one were to use this HPDI, one could conclude that permanent shocks either play a negligible role in explaining fluctuations in wealth or they play the dominant role! Interestingly, it is the Shrinkage prior approach, which is the most “subjective” one we adopt (in the sense that we need to select a prior for the shrinkage parameter), which yields the most regular and well-behaved results.

A key conclusion from Table 4 is that it is hard to make a definitive conclusion about the relative role of transitory and permanent shocks in driving fluctuations in wealth (or any of the variables). We are presenting intervals which contain 50% of the posterior probability (rather than more usual 95%). But even these are very wide, indicating that the data is not very informative about the relative roles of permanent and transitory shocks. A related conclusion is that presenting point estimates such as posterior means can be very misleading. Indeed there are many cases where the posterior mean is not even included in the HPDI! The point estimates in Table 4 are not that dissimilar to those presented in LL (see their Table 2). In general, these indicate (as found in LL) that permanent shocks have a relatively small role to play in driving fluctuations in wealth, but have a larger role in driving fluctuations in income and consumption. However, given the multimodal nature of the posterior and the fact that appreciable probability is allocated to regions inconsistent with this story, it is risky indeed to place much store in point estimates.

This finding can be seen more clearly in Figure 1 which, for the wealth variable, plots the entire posterior distributions which were used to create Table 4. All of the posteriors have a substantial mode in LL’s region, but non-negligible weight is allocated to all regions of the parameter space. Furthermore, with the exception of the Shrinkage prior approach, there is an appreciable mode at 1 (i.e. the point where the permanent component is completely dominant – the point exactly opposite of LL’s story). To shed more insight on the properties of the posterior in Figure 1, note that the probabilities allocated to the region  $[0, 0.5]$  (i.e. the region where permanent shocks are less important than transitory shocks in the variance decomposition) are, respectively, 0.575, 0.739 and 0.859 for the BIC, fractional Bayes factor and Shrinkage prior approaches. Thus, although there is appreciable support for the view that transitory shocks dominate fluctuations in wealth, this support

is far from being overwhelming.

For the consumption and income variables, the posteriors also exhibit substantial dispersion and multimodality. However, with the exception of the fractional Bayes Factor approach, there is more of a consensus that it is the permanent shocks that are playing the predominant role. Such a finding is consistent with LL.

Table 4: Variance Decompositions Using All Models Share of Forecast Error Due to Permanent Component			
	Wealth	Consumption	Income
BIC			
Mean	0.496	0.821	0.823
Median	0.307	0.750	0.750
50% HPDI	[0.021, 0.136] [0.964, 1.000]	[0.736, 0.764] [0.779, 1.000]	[0.750, 1.000]
IQR	[0.079, 0.821]	[0.436, 0.936]	[0.436, 0.936]
Fractional Bayes Factor			
Mean	0.236	0.354	0.411
Median	0.207	0.436	0.636
50% HPDI	[0.000, 0.136] [0.279, 0.293] [0.464, 0.479]	[0.000, 0.036] [0.150, 0.164] [1.000, 1.000]	[0.000, 0.036] [0.050, 0.064] [0.364, 0.379] [0.621, 0.636] [0.679, 0.707] [1.000, 1.000]
IQR	[0.064, 0.521]	[0.064, 1.000]	[0.293, 0.593]
Shrinkage Prior			
Mean	0.235	0.771	0.827
Median	0.150	1.000	1.000
50% HPDI	[0.079, 0.193]	[1.000, 1.000]	[1.000, 1.000]
IQR	[0.093, 0.221]	[0.479, 1.000]	[0.693, 1.000]
Average over all Approaches			
Mean	0.332	0.602	0.628
Median	0.136	0.636	0.679
50% HPDI	[0.000, 0.136] [1.000, 1.000]	[0.000, 0.050] [0.064, 0.093] [0.450, 0.464] [0.921, 0.936] [0.979, 1.000]	[0.000, 0.050] [0.079, 0.093] [0.121, 0.136] [0.264, 0.293] [0.379, 0.393] [0.450, 0.464] [0.950, 1.000]
IQR	[0.064, 0.650]	[0.136, 1.000]	[0.279, 1.000]

Table 5 presents the same information as Table 4, except that we average only over models consistent with economic theory. Figure 2, analogous to Figure 1, plots the posterior of a particular variance decomposition of interest. It can be seen that imposing the restrictions of economic theory only slightly alters the posterior distributions of the variance decompositions. Even when we are ruling out  $r = 0$  or  $3$ , we still find that models with  $r = 1$  or  $2$  still attach appreciable support to regions of the parameter space which are inconsistent with LL's story. For instance, they allow for substantial probability that it is permanent shocks to wealth which are

playing the dominant role. *A priori*, one might have expected that the inclusion/exclusion of models with  $r = 0$  (which imply all shocks are permanent) would be crucial. Table 5 indicates that this is not so. In fact, using the fractional Bayes factor approach, imposing economic theory actually implies slightly *less* support for the "transitory shocks are predominant in driving wealth" story (e.g. the probability in the interval  $[0, 0.5]$  drops from 0.739 to 0.642 for this case). This arises since we are also ruling out  $r = 3$  (which implies all shocks are transitory),

Table 5: Variance Decompositions Models Consistent with Theory Share of Forecast Error Due to Permanent Component			
	Wealth	Consumption	Income
BIC			
Mean	0.410	0.808	0.809
50% HPDI	[0.021, 0.136] [0.979, 0.1.000]	[0.754, 0.764] [0.779, 0.1.000]	[0.750, 1.000]
Fractional Bayes Factor			
Mean	0.374	0.345	0.809
50% HPDI	[0.000, 0.136] [0.164, 0.179] [0.521, 536] [0.792, 0.807] [0.879, 0.893] [0.936, 0.950]	[0.000, 0.036] [1.000, 1.000]	[0.000, 0.036] [0.064, 0.093] [0.236, 0.250] [0.279, 0.293] [0.464, 0.479] [0.535, 0.550] [0.636, 0.650] [0.679, 0.693] [1.000, 1.000]
Shrinkage Prior			
Mean	0.239	0.757	0.817
50% HPDI	[0.079, 0.193]	[1.000, 1.000]	[1.000, 1.000]
Average over All Approaches			
Mean	0.369	0.715	0.761
50% HPDI	[0.021, 0.150] [1.000, 1.000]	[0.000, 0.079] [0.951, 1.000]	[0.000, 0.036] [0.050, 0.064] [0.279, 0.293] [0.507, 0.521] [0.550, 0.564] [0.579, 0.592] [0.650, 0.664] [0.721, 0.736] [0.864, 1.000]

To provide evidence that the conflict between our results and LL's is due to our treatment of model uncertainty, and not due to the use of Bayesian methods within a given model, Table 6 presents results for the model used by LL with  $r = 1, d = 3$  and  $p = 1$ . We average across models with and without the weak exogeneity restriction imposed. For  $r = 1, d = 3$  and  $p = 1$ , the probability of weak exogeneity holding is quite high for all approaches (i.e. 0.981, 0.789 and 1.000 for our three approaches). Thus, results in Table 6 are quite close to those we found using the single model  $r = 1, d = 3$  and  $p = 1$  with weak exogeneity imposed (i.e. the model which LL favor).



For the sake of brevity, we only present results for the wealth variable. For comparison, LL present (see their Table 2) a point estimate of this variance decomposition of 0.119 and a 95% confidence interval of [0.11, 0.20]. Table 6 indicates a very slightly lower variance decomposition, but overall our results are quite similar. In short, in the context of this model, we are recovering LL’s findings that only a very small part of the fluctuations in wealth are driven by permanent innovations and that this result is fairly precisely estimated.

	BIC	Fractional Bayes Fact.	Shrinkage Prior	Average over All Approaches
Mean	0.075	0.075	0.074	0.064
Median	0.064	0.064	0.064	0.064
50% HPDI	[0.050, 0.093]	[0.050, 0.093]	[0.050, 0.093]	[0.050, 0.093]
IQR	[0.050, 0.093]	[0.050, 0.093]	[0.050, 0.093]	[0.050, 0.093]

A deeper investigation into exactly which of LL’s specification choices are crucial is revealed in Figure 3. This figure plots, for the wealth variable, the posteriors of the share of the forecast error variance due to the permanent component for LL’s model ( $r = 1, d = 3$  and  $p = 1$ ) without exogeneity imposed and averaged over the models with and without exogeneity. For brevity, we only plot the posterior averaged over our three approaches. The posterior which averages over models with and without exogeneity is well-behaved (this is the posterior which underlies the results in Table 6). However, when we do not impose weak exogeneity, then the posterior alters dramatically. In particular, we now observe a much flatter posterior which allocates more weight near the point most inconsistent with LL’s story. Clearly imposition of weak exogeneity is having a huge effect on the posterior.<sup>15</sup>

Before completing the empirical work, we had thought that  $r$  would be the key model aspect that was important for results. After all, models with  $r = 0$  imply that all shocks are permanent. Surely this aspect could account for why graphs like Figure 1 had a spike of probability near 1. However, although choice of cointegrating rank is important, it turns out that the imposition of weak exogeneity is even more important. After all, Figure 2 rules out the case  $r = 0$  and it still exhibits many of the properties of Figure 1. Even the single model of LL, without weak exogeneity imposed, exhibits some such properties. It is only when we set  $r = 1$  and impose weak exogeneity do we obtain results that are fully consistent with the story that permanent shocks have

<sup>15</sup>It is interesting to note that, when LL do not impose weak exogeneity, they too are finding a higher point estimate with a much wider confidence interval.

little role to play in driving wealth. An examination of the formulae underlying the permanent/transitory split and variance decomposition offers a possible explanation as to why weak exogeneity might be so crucial. The variance decompositions are complicated nonlinear transformations of the VECM parameters. In particular, the parameters relating to weak exogeneity (i.e.  $\gamma$ ) often appear in the denominators of fractions determining the long run effects of shocks. If these are imprecisely estimated, then they can have a substantial impact on the variance decompositions. Potter (2001) offers a detailed discussion of this and related points.

It would be desirable to consider every one of our models with and without this weak exogeneity restriction imposed. However, as discussed in the previous section, the particular weak exogeneity restriction used by LL is only possible in models with  $r = 1$ . So a strategy of simply doubling the number of models under consideration (with/without weak exogeneity imposed) is not possible. It would be possible to work with our existing set of models plus an additional 20 models with  $r = 1$  with weak exogeneity imposed. If we do this, our results are pulled more towards those of LL. However, since appreciable posterior weight is attached to models without LL's weak exogeneity restrictions imposed (i.e. for  $r = 0, 2, 3$ ), we still obtain posteriors of our variance decompositions which are quite dispersed, often multi-modal and attach at least some weight at boundaries. In short, it is hard to see how we can precisely estimate variance decompositions in this data set without making strong assumptions that are not fully supported by the data.

As one final piece of evidence relating to the importance of weak exogeneity, we obtained variance decompositions averaged over the set of models consistent with economic theory (i.e.  $d > 2$  and  $r = 1$  and 2) with weak exogeneity restrictions imposed throughout. For models with  $r = 1$  we impose the same weak exogeneity restriction as above. For  $r = 2$  we impose a weak exogeneity restriction where only consumption is weakly exogenous in the cointegrating relationship. This can be interpreted as imposing a form of the permanent income hypothesis and there is appreciable support for this hypothesis in the data set (e.g. for models with  $r = 2$  and  $d = 3$  the probability of this restriction holding varies from 0.674 to 1.000 in our various approaches and lag lengths). Thus, for  $r = 1$  and 2, we are imposing different weak exogeneity restrictions (which some econometricians may object to). When we do this, we recover results quite similar to LL's. For instance, the posterior relating to the share of the permanent component in wealth fluctuation (i.e. analogous to the one plotted in the figures) is well-behaved with posterior mean being 0.110 and 50% HPDI being [0.050, 0.136]. That is, if we impose something like permanent income theory on the behavior of consumption (either in the case of

$r = 1$  or 2) most of the adjustment to transitory shocks comes through the wealth variable

There are many other features that an empirical researcher may be interested in (e.g. other sorts of variance decompositions, long-horizon regressions, etc.).<sup>16</sup> However, for the sake of brevity, we end our empirical results here, having established our basic point: that model uncertainty is an important issue in this data set and that empirical conclusions depend on the way this issue is treated.

## 4 Conclusions

In this paper we have re-examined evidence relating to the consumption-wealth relationship using a variety of different approaches to Bayesian model averaging. We document a great deal of uncertainty with regards to the number of cointegrating vectors, lag length, form of deterministic trends in a commonly-used data set. As a result of this, we find that the exact magnitude of the role of permanent shocks is hard to estimate precisely. Thus, although some support exists for the view that their role is small, we cannot rule out the possibility that they have a substantive role to play. In particular, the posterior distributions of key variance decompositions are found to be multimodal and relatively non-informative. It is only if we work with the single model used by LL that we obtain results very similar to those in LL. Within that single model the restriction that the cointegrating residual only enters the wealth equation used by LL turns out to be of great importance in producing the result that most fluctuations in wealth are transitory.

In addition to making a contribution to the empirical literature on the wealth effect, we have presented a broader methodological argument that the conclusions of any empirical exercise should depend on economic theory, the data and the treatment of model uncertainty. We view Bayesian methods as a tool that allows us to investigate the relative roles of these three aspects. That is, Bayesian methods allow us to separate *a priori* beliefs about economic theory from the information in the observed data and their data-driven model specification choices.

## 5 References

Brennan, M. and Xia, Y. (2004) “Tay’s as good as cay,” *Finance Research Letters*, forthcoming.

---

<sup>16</sup>One of the strengths of LL’s original approach is the fact that the cay variable is useful for predicting asset returns. There is some considerable controversy over the source of this predictability due to the fact that the cay is estimated over the whole sample [see Brennan and Xia (2004)].

- Centoni, M. and Cubadda, G. (2003). "Measuring the business cycle effects of permanent and transitory shocks in cointegrated time series," *Economics Letters*, 80, 45-51.
- Chikuse, Y. (1990). "The matrix angular central Gaussian distribution," *Journal of Multivariate Analysis* 33, 265-274.
- Cogley, T. and Sargent, T. (2003). "Drifts and volatilities: monetary policies and outcomes on Post WWII US," working paper.
- Doan, T., Litterman, R. and Sims, C. (1984). "Forecasting and conditional projection using realistic prior distributions," *Econometric Reviews*, 3,1-100.
- Draper, D. (1995). "Assessment and propagation of model uncertainty (with discussion)," *Journal of the Royal Statistical Society Series B*, 56, 45-98.
- Fernández, C., Ley, E. and Steel, M. (2001). "Model uncertainty in cross-country growth regressions," *Journal of Applied Econometrics*, 16, 563-576.
- Garratt, A., Lee, K., Pesaran, M.H., Shin, Y. (2003). "Forecast uncertainties in macroeconomic modelling: An application to the UK economy," *Journal of the American Statistical Association*, 98, 829-838.
- Gelfand, A.E., Dey, D.K. (1994). "Bayesian model choice: asymptotics and exact calculations," *Journal of the Royal Statistical Society, Series B* 56, 501-504.
- Gonzalo, J. and Ng, S. (2001). "A systematic framework for analyzing the dynamic effects of permanent and transitory shocks," *Journal of Economic Dynamics and Control*, 25, 1527-1546.
- Hjort, N. and Claeskens, G. (2003). "Frequentist model average estimators," *Journal of the American Statistical Association*, 98, 879-899.
- Hodges, J. (1987). "Uncertainty, policy analysis and statistics," *Statistical Science*, 2, 259-291.
- Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.
- Julliard, C. (2005). Labor income and wealth, Department of Economics, Princeton University, manuscript.
- Kass, R. and Raftery, A. (1995). "Bayes Factors," *Journal of the American Statistical Association*, 90, 773-795.
- Kleibergen, F. and Paap, R. (2002). "Prior, posteriors and Bayes factors for a Bayesian analysis of cointegration," *Journal of Econometrics*, 111, 223-249.

- Koop, G., Strachan, R., van Dijk, H. and Villani, M. (2004). "Bayesian approaches to cointegration," prepared for T.C. Mills and K. Patterson (eds.) Palgrave Handbook of Econometrics: Volume 1 Econometric Theory, Basingstoke, Palgrave Macmillan, 2005 (forthcoming). Available at <http://www.le.ac.uk/economics/gmk6/>.
- Leamer, E. (1978). *Specification Searches*. New York: Wiley.
- Lee, H. (2002). "Deterministic trend and biases in the estimated consumption-wealth ratio," working paper available at <http://www.columbia.edu/~hl306/cay.pdf>.
- Lettau, M. and Ludvigson, S. (2004). "Understanding trend and cycle in asset values: Reevaluating the wealth effect on consumption," *American Economic Review*, 94, 276-299.
- Lettau, M. and Ludvigson, S. (2001). "Consumption, aggregate wealth and expected stock returns," *Journal of Finance*, 56, 815-849.
- Min, C. and Zellner, A. (1993). "Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates," *Journal of Econometrics*, 56, 89-118.
- O'Hagan, A. (1995). "Fractional Bayes Factors for model comparison," *Journal of the Royal Statistical Society, Series B*, 57, 99-138.
- Poirier, D. (1995). *Intermediate Statistics and Econometrics: A Comparative Approach*. Cambridge: The MIT Press.
- Potter, S. (2001). "Using *a priori* information in the analysis of cointegrated systems," working paper.
- Raftery, A., Madigan, D. and Hoeting, J. (1997). "Bayesian model averaging for linear regression models," *Journal of the American Statistical Association*, 92, 179-191.
- Rudd, J. and Whelan, K. (2002). "A note on the cointegration of consumption, income and wealth," working paper available at <http://www.federalreserve.gov/pubs/feds/2002/200253/200253pap.pdf>.
- Sala-i-Martin, X. (1997). "I just ran two million regressions," *American Economic Review*, 87, 178-183.
- Sala-i-Martin, X., Doppelhoffer, G and Miller, R. (2004). "Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach," *American Economic Review*, 94, 813-835.
- Schwarz, G. (1978). "Estimating the dimension of a model," *Annals of Statistics* 6:2, 461-464.
- Strachan, R. W. (2003). "Valid Bayesian estimation of the cointegrating error correction model," *Journal of Business and Economic Statistics*, 21, 185-195.

Strachan, R. W. and Inder, B. (2004). "Bayesian analysis of the error correction model," *Journal of Econometrics*, forthcoming.

Strachan, R. W. and van Dijk, H. K. (2004). "The value of structural information in the VAR model," Econometric Institute Report EI 2004-23, Erasmus University Rotterdam.

Villani, M. (2004). "Bayes reference analysis of cointegration," *Econometric Theory*, forthcoming.

## 6 Appendix: Bayesian Econometric Methods

The posterior probabilities for each model,  $p(M_r|Data)$ , are required for the model averaging exercises. Estimates of these probabilities are calculated from estimates of the marginal likelihoods,  $m_r(y)$ , via the expression

$$p(M_r|Data) = \frac{m_r(y)}{\sum_i m_i(y)}. \quad (A.1)$$

We have summed over all models in the denominator in (A.1). In this paper, we estimate the marginal likelihoods using three methods. First, we adopt the fractional Bayes factor approach of O'Hagan (1995). Second, the Bayesian information criteria (BIC) of Schwarz (1978). Finally, we use Markov Chain Monte Carlo (MCMC) in the context of a certain reference prior described below. Here we provide a brief description of the first two which provide asymptotic approximations to  $m_r(y)$  without requiring specification of a prior distribution for the parameters. The remainder of the appendix outlines the third approach.

The fractional Bayes factor is a particular form of partial Bayes factor proposed by O'Hagan (1995). As a means of obtaining estimates of Bayes factors with weak or improper priors, the partial Bayes factor uses a fraction of the sample - or a training sample - to produce a posterior which is then used as a prior for the remainder of the sample. One arbitrary feature of this method is the choice of particular subsample to use as the training sample. O'Hagan proposed the fractional Bayes factor to overcome this difficulty and so reduce the problem to simply having to choose 'how much', or what fraction,  $b = i/T$ , of the sample to use in the training sample. In our empirical work, we use  $b = 14/201$ . Having made a selection as to the fraction of the sample to use in the training sample, we can then use O'Hagan's expression (14) for an asymptotic approximation to  $-2$  times the log Bayes factor, to obtain an expression for  $-2$  times the log of the marginal likelihood for a particular model as

$$-2 \ln m_r(y) \approx (1 - b) \left\{ -2\widehat{l}_r - p_r \ln \frac{b}{(1 - b)} \right\},$$

where  $\widehat{l}_r$  is the maximized log likelihood for model  $M_r$  and  $p_r$  is the number of parameters in the model. This gives us a simple expression from which we can obtain an estimator for the Bayes factors and then the posterior model probabilities via (A.1).

Schwarz (1978) presents an asymptotic approximation to the marginal likelihood of the form

$$\ln m_r(y) \approx \widehat{l}_r - \frac{p_r \ln T}{2}.$$

This is  $-2/T$  times the well known BIC or SBC commonly used for model selection. As with the fractional Bayes factor, we can obtain an estimate of  $p(M_r|Data)$  from this expression via (A.1). These techniques, which are based upon asymptotic arguments, have the advantages of consistency and simplicity. Furthermore, subjective elicitation of prior hyperparameters is not required.

The third method we use is an approximation to  $m(y)$  based upon analytical integration in some dimensions and a MCMC sampling technique in other dimensions where analytical results are not available. We briefly introduce the approach here and provide more details in the following subsections. First, we collect the parameters in the model  $M_r$  into  $(\Sigma, \theta, \alpha)$  where  $\theta$  is a vector containing the elements in  $\gamma$ ,  $\mu$ , and  $\Gamma(L)$  and  $\Sigma$  is the covariance matrix for the multivariate Normal errors,  $\varepsilon_t$ . Next, an improper diffuse prior is specified for  $\Sigma$ , a relatively diffuse but proper prior is specified for  $\theta$  and a flat (proper) prior is specified for  $\alpha$ . This defines our prior  $p(\Sigma, \theta, \alpha)$ . Combining the prior with the likelihood,  $L(\Sigma, \theta, \alpha)$ , we obtain an expression proportional to the posterior,  $p(\Sigma, \theta, \alpha|M_r, Data)$ , such that

$$m_r(y) = \int \int p(\Sigma, \theta, \alpha) L(\theta, \alpha) d\alpha d(\Sigma, \theta).$$

As a closed form expression (conditional upon  $\alpha$ ) exists for the outer integral with respect to  $(\Sigma, \theta)$ , we can perform this integral analytically. It is the inner integral with respect to  $\alpha$  for which we have no closed form expression and so resort to MCMC methods. We use the approach proposed in Strachan and Inder (2004) and Strachan and van Dijk (2004) in which a Uniform prior is placed upon the (cointegrating) space of  $\alpha$  and  $\alpha$  is specified as a semiorthogonal matrix such that  $\alpha'\alpha = I_r$ . This specification ensures the posterior will be proper for all models considered and a simple, efficient sampling scheme can be specified to draw  $\alpha$ . Further details are provided below and in the survey paper Koop, Strachan, van Dijk and Villani (2004).

## 6.1 The Prior

We use the following commonly used noninformative prior for  $\Sigma$

$$p(\Sigma) \propto |\Sigma|^{-(n+1)/2}. \quad (\text{A.2})$$

This is an improper prior but will cause no problem for computation of the Bayes factor as this same prior is employed for all models. Kass and Raftery (1995) cite a number of authors who accept this principle and Fernández, Ley and Steel (2001) employ this procedure and the same prior in a model averaging exercise. As Fernández et al. point out, the prior (A.2) is invariant to scale transformations and, although it is not strictly Jeffreys' prior, it is that part of Jeffreys' prior related to  $\Sigma$  and widely accepted as a non-informative prior for  $\Sigma$ .

We next specify a Normal prior for  $\theta$  as  $N(0, \Sigma \otimes I\nu)$  where  $\nu$  is a hyperparameter in the variance which controls how diffuse is the prior. It is this hyperparameter  $\nu$  which motivates our terminology "shrinkage" prior. As we did not wish to select a particular value for  $\nu$  but wanted to keep the prior for  $\theta$  diffuse, we gave  $\nu$  a Gamma distribution  $g(n_1, n_2)$  in which moderately diffuse values for  $n_1$  and  $n_2$  are chosen ( $n_1 = 3, n_2 = 4$ ). This prior places 99% of the prior mass for  $\nu$  in the interval (0.45, 19.9), ensuring the prior for  $\theta$  will cover a wide range of plausible values.

The prior for  $\alpha$  is implied by a Uniform prior over the cointegration space. As is frequently stated in empirical cointegration studies, the elements of  $\alpha$  are not identified and so restrictions need to be imposed to permit their estimation. The only information we can obtain from likelihood based analysis of a VECM is information on the location of the cointegrating space, or the space spanned by  $\alpha$ . As argued in Villani (2004) and Strachan and Inder (2004), this is not a limitation since it is the cointegrating space that is the object of interest and estimates of  $\alpha$  simply provide an estimate of this space. We wish to be non-informative about the cointegrating space and so follow the approach in Strachan and Inder (2004) by specifying  $\alpha$  to be semiorthogonal such that  $\alpha'\alpha = I_r$  so  $\alpha$  has a compact support, and place a Uniform prior on this support. Alternative specifications may be used to express ignorance about the cointegrating space, but this is the only one proposed to date that ensures we can obtain proper posterior distributions for all of the models we consider.



## 6.2 Posterior Computation

The prior specification above is also used in Strachan and van Dijk (2004) for a range of model averaging exercises. They show that the marginal density for  $\alpha$ ,  $d$ ,  $p$  and  $r$  is

$$p(\alpha, d, p, r|y) \propto \mathfrak{g}_{d,p,r} k(\alpha)$$

where  $\mathfrak{g}_{d,p,r}$  is a function of the data and  $d$ ,  $p$  and  $r$  and

$$k(\alpha) = |\alpha' D_0 \alpha|^{-T/2} |\alpha' D_1 \alpha|^{(T-n)/2}.$$

where  $D_0$  and  $D_1$  are data-based quantities [see Strachan and van Dijk (2004) for precise definitions]. The integral  $m = \int k(\alpha) d\alpha$  does not have a closed form, therefore to obtain the marginal likelihood for a model we use MCMC methods to calculate  $m$ . To do this, we require both a series of draws of  $\alpha$  from its posterior distribution,  $p(\alpha|d, p, r, y)$ , and a method of using these draws to estimate  $m$ . We use a random walk Metropolis-Hastings algorithm to obtain 80,000 draws of  $\alpha$  from  $p(\alpha|d, p, r, y)$  (after a burn-in of 10,000 draws) in which the candidate distribution is the matrix angular central Gaussian distribution,  $G(\alpha)$ , with density  $g(\alpha)$ . This distribution was first proposed by Chikuse (1990) and the derivation of the specific form used in this paper is the same as that given in Strachan and Inder (2004). To obtain a draw from this candidate involves the following steps. Assume we have a current draw of  $\alpha$  and therefore the  $n \times (n-r)$  matrix lying in the orthogonal complement of  $\alpha$ ,  $\alpha_{\perp}$ . Denote these draws respectively as  $\alpha^{(i)}$  and  $\alpha_{\perp}^{(i)}$ . Then iterate over the following steps.

- Draw  $\tau$  from  $N(0, \sigma)$ .
- Construct  $P = \alpha^{(i)} \alpha^{(i)'} + \alpha_{\perp}^{(i)} \alpha_{\perp}^{(i)'} \tau$ .
- Draw an  $n \times r$  matrix  $Z = \{z_{ij}\}$  from the Multivariate Standard Normal such that each  $z_{ij} \sim N(0, 1)$  and independent of all other  $z_{ij}$ .
- Construct  $X = PZ$ .
- Decompose  $X$  into  $X = \alpha^* \kappa$  where  $\alpha^{*'} \alpha^* = I_r$  and  $\kappa$  is full rank upper triangular.
- The matrix  $\alpha^*$  is a candidate draw from  $G(\alpha)$  which is 'located' at the previous value  $\alpha^{(i)}$ .

- Accept  $\alpha^*$  and set  $\alpha^{(i+1)} = \alpha^*$  with probability  $\min \left\{ \frac{p(\alpha^*)g(\alpha^{(i)})}{p(\alpha^{(i)})g(\alpha^*)}, 1 \right\}$ , else set  $\alpha^{(i+1)} = \alpha^{(i)}$ .

The only parameter value to be decided for  $g(\alpha)$  is  $\sigma$ . Suggestions for how to choose this value are given in Strachan and Inder (2004). We use  $\sigma = 0.6$  as we find this keeps the tails thin but still allows reasonable dispersion.

To estimate  $m$  we take the approach proposed by Gelfand and Dey (1994) and use a proper distribution on the support of  $\alpha$ , draws from the posterior for  $\alpha$  and the posterior kernel  $k(\alpha)$ . If  $g(\alpha)$  is a proper density, we may obtain the estimate of  $m$  from

$$\frac{1}{m} = \int \frac{g(\alpha) k(\alpha)}{k(\alpha) m} d\alpha.$$

As we have a sequence of draws  $\alpha^{(i)}$ ,  $i = 1, \dots, J$ , from the distribution with density  $k(\alpha)/m$ , we can estimate  $m$  by

$$\hat{m} = J \left( \sum_{i=1}^J \frac{g(\alpha^{(i)})}{k(\alpha^{(i)})} \right)^{-1}.$$

Our choice of  $g(\alpha)$ , as the notation suggests, is again the matrix angular central Gaussian distribution,  $G(\alpha)$ , with density  $g(\alpha)$ . However, the location  $\alpha^\bullet$  is fixed at the value of  $\alpha^{(i)}$  that gave the highest value of  $g(\alpha)$  over the burn-in sample. All computations were performed using Gauss 3.5.

A different approach was taken to estimate the posterior probabilities for the models with weak exogeneity imposed. We use the draws  $\alpha^{(i)}$  from the distribution with kernel  $k(\alpha)$  for the model without weak exogeneity imposed and denote the kernel for the same model with weak exogeneity imposed as  $k_x(\alpha)$ . The marginal likelihoods for these models are respectively  $m = \int k(\alpha) d\alpha$  and  $m_x = \int k_x(\alpha) d\alpha$  and the Bayes factor

$$B_x = \frac{m_x}{m} = \int \frac{k_x(\alpha) k(\alpha)}{k(\alpha) m} d\alpha$$

can then be estimated as

$$\hat{B}_x = \frac{1}{J} \sum_{i=1}^J \frac{k_x(\alpha^{(i)})}{k(\alpha^{(i)})}.$$

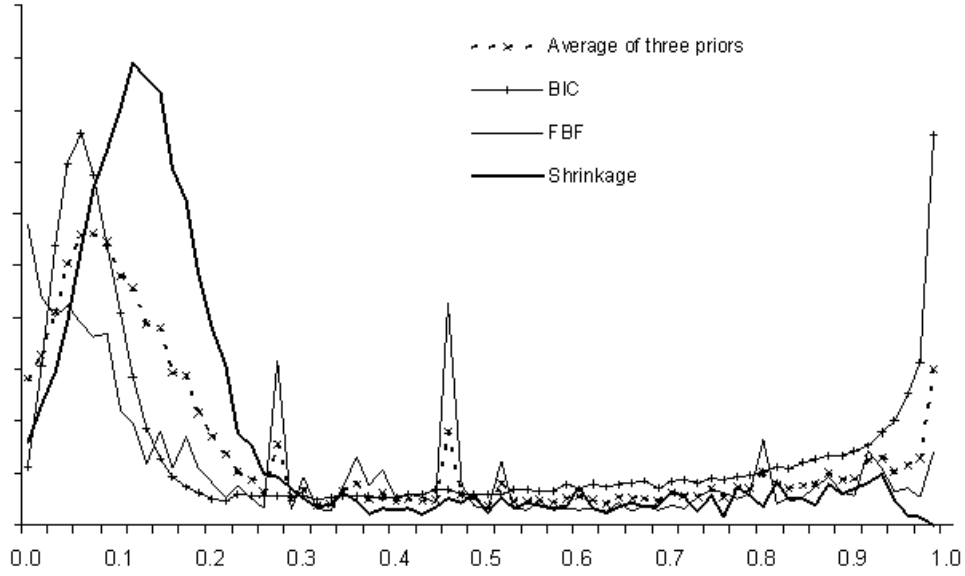


Figure 1: Posterior of Share of Forecast Error Variance in Wealth Due to Permanent Component (averaged across all models).

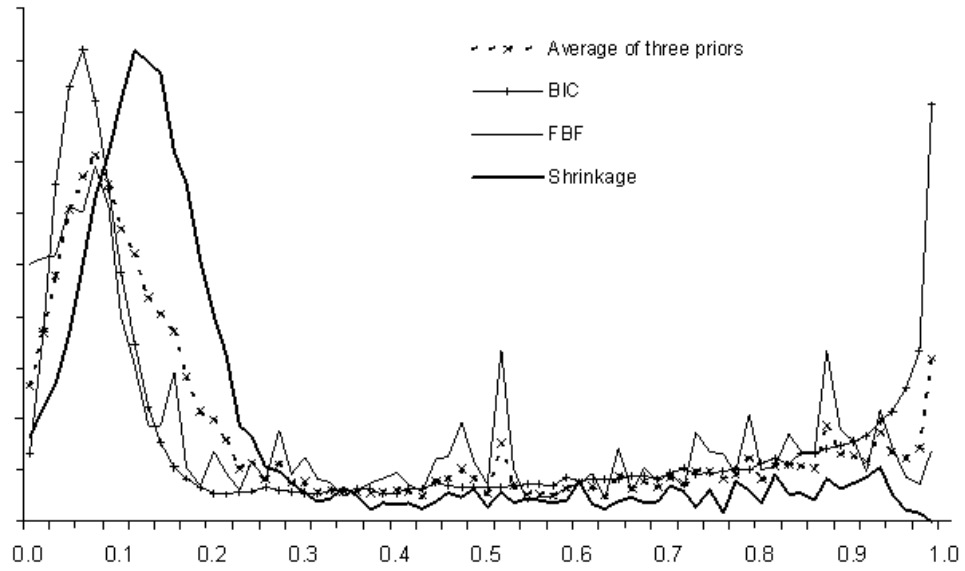


Figure 2: Posterior of Share of Forecast Error Variance in Wealth Due to Permanent Component (averaged across models consistent with economic theory).

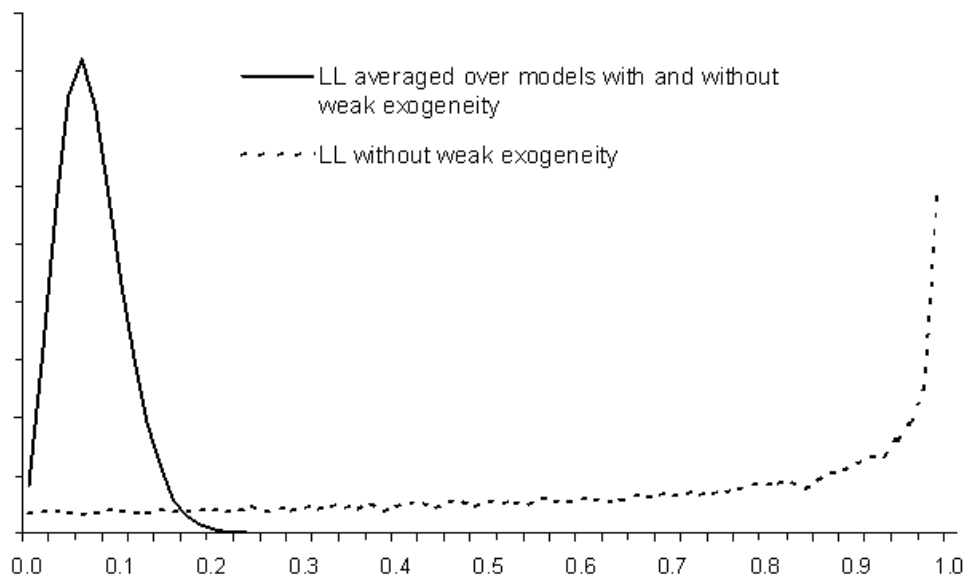


Figure 3: Posterior of Share of Forecast Error Variance in Wealth Due to Permanent Component (for the model specification of Lettau and Ludvigson - LL). The solid line shows the pdf for the LL specification averaged over models with and without weak exogeneity. The dashed line shows the pdf for the model without weak exogeneity.