GARY KOOP
SIMON M. POTTER
RODNEY W. STRACHAN

Re-Examining the Consumption–Wealth Relationship: The Role of Model Uncertainty

This paper discusses the consumption–wealth relationship. We use data on consumption, assets, and labor income and a vector error correction framework. This framework defines a set of models that differ in the number of co-integrating vectors, the form of deterministic components and lag length. Further models can be defined through parametric restrictions and, in particular, interest centers on a weak exogeneity restriction that says that the co-integrating residuals do not affect consumption and income directly. Key results in previous work relate to the roles of permanent and transitory shocks in driving wealth and how consumption responds to these shocks. We investigate the robustness of these results to model uncertainty and argue for the use of Bayesian model averaging. We find that there is a large degree of model uncertainty. Whether this uncertainty has important empirical implications depends on the researcher's attitude toward the theory used to motivate a co-integrating relationship between consumption, assets and income. If we work with models consistent with this theory and impose the weak exogeneity restriction, we find precisely estimated results that show that permanent shocks have only a small role in driving assets and that the predominant transitory shocks have little effect on consumption. These findings are consistent with the previous literature. However, if we work with a broader set of models and let the data speak, 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.

JEL codes: C11, C22, E21
Keywords: wealth effect, vector error correction model, Bayesian model averaging, co-integration, variance decomposition.

The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of New York or the Federal Reserve System. We would like to thank Sydney Ludvigson and seminar participants at the University of Birmingham, Orebro University, University of Exeter and the Institute for Advanced Studies (Vienna), for helpful comments.

GARY KOOP *is from the Department of Economics, University of Strathclyde* (*E-mail:* Gary.Koop@strath.ac.uk). SIMON M. POTTER *is from the Federal Reserve Bank of New York* (*E-mail:* Simon.Potter@ny.frb.org). RODNEY W. STRACHAN *is from the School of Economics, University of Queensland* (*E-mail:* r.strachan@uq.edu.au).

Received August 5, 2005; and accepted in revised form April 23, 2007.

Journal of Money, Credit and Banking, Vol. 40, No. 2–3 (March–April 2008)

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. However, recent events that 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 that attempts to reconcile the apparent conflict between textbook wisdom and recent experience. Using co-integration 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. This is an attractive and influential story. Other authors (e.g., among others, Brennan and Xia 2005, Lee 2002, Palumbo, Rudd, and Whelan 2006, Rudd and Whelan 2006) have adopted similar econometric methods, not necessarily finding the same story. Rudd and Whelan (2006), for instance, do not even find co-integration (using different data definitions) and, thus, must conclude that all shocks to wealth are permanent.

Like many empirical conclusions in economics, the conclusions in this literature are usually 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 that treats the final selected model as though it were the truth 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 (2005), 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.[1] The basic idea of Bayesian model averaging can be explained quite simply. Suppose the researcher is entertaining R possible models, denoted by $M_1, \ldots, M_R$, to learn about an unknown feature of the economy, ϕ (e.g., a variance decomposition or an impulse response to a particular shock). If we treat ϕ and M_r as random variables, the rules of conditional expectation imply that:

1. Papers that investigate the use of non-Bayesian model averaging techniques in macro-economic applications include Garratt et al. (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).

$$E(\phi \mid Data) = \sum_{r=1}^{R} p(M_r \mid Data) E(\phi \mid Data, M_r). \tag{1}$$

Thus, overall point estimates, $E(\phi \mid Data)$, should be an average of point estimates in individual models, $E(\phi \mid Data, M_r)$.[2] The weights in the average are the posterior model probabilities, $p(M_r \mid Data)$. 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, pp. 519–523) to formal justifications of such an approach in decision theoretic contexts (e.g., Min and Zellner 1993, Raftery, Madigan, and Hoeting 1997).

Given the importance for macro-economic policy of results relating to the consumption–wealth ratio and the large and growing literature using similar specifications (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 methods for co-integrated models, we carry out a Bayesian model averaging exercise where different models are defined depending on the number of co-integrating vectors, form of deterministic trends, lag length and a weak exogeneity restriction (i.e., whether the co-integrating residuals affect consumption and income directly). We find that many different models receive appreciable support from the data and, thus, substantial model uncertainty exists. However, just because substantial model uncertainty exists does not necessarily mean there will be substantial uncertainty about an economic feature of interest (ϕ). For instance, it is possible that all our models yield similar variance decompositions and, thus, that model uncertainty has no economic implications. However, in our empirical work, we do find that model uncertainty matters. In particular, we find that the exact magnitude of the role of permanent shocks in wealth 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 multi-modal and relatively non-informative. In contrast, when we consider only a single model consistent with theoretical considerations and impose a weak exogeneity restriction that says that the co-integrating residual only directly influences wealth (and not consumption or income), we find precisely estimated variance decompositions that tell a strong story that permanent shocks have a small role in driving wealth. These latter results are consistent with previous findings in the literature and, in particular, with LL. We do not argue that one set of results are "correct" and the other "incorrect." Instead, we argue that it is important to be clear about the role of the data, theoretical considerations (e.g., should models consistent with economic theory be viewed, *a priori*, as more plausible than other models?) and econometric methods in producing empirical results. Our Bayesian model averaging exercise offers a clear way of disentangling these roles.

2. The same argument holds for other features of the posterior distribution (e.g., predictives or measures of dispersion such as Highest Posterior Density Intervals).

1. CO-INTEGRATION AND THE CONSUMPTION–WEALTH RELATIONSHIP

Many studies of the consumption–wealth relationship begin with 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. If we log-linearize the budget constraint relationship, assume consumption growth and the returns on human capital and asset wealth are stationary and relate the non-stationary component of human capital to labor income (see, e.g., Lettau and Ludvigson 2001), the result is a model where log consumption, c_t, log asset wealth, a_t, and log labor income, y_t, should be co-integrated. Often empirical studies use a single co-integrating relationship referred to as "cay," $c_t - \alpha_a a_t - \alpha_y y_t$. That is, the co-integrating vector should be $(1, -\alpha_a, -\alpha_y)'$ and there should be no deterministic trend in the co-integrating residual (although there may be an intercept). This implies that there are two permanent shocks and one transitory shock driving the joint fluctuations in consumption, labor income and wealth. This accounting argument does not rule out the case of two co-integrating relationships, with the implication of one permanent and two transitory shocks and once again no deterministic trends in the co-integrating residuals. Note that in this latter case "cay" is still stationary.

An important methodological issue is how to include the theoretical information outlined in the previous paragraph in a class of statistical models. Before we directly address this issue, let us define the class of statistical models. In this paper, as in much of the related research (such as LL), Vector Error Correction models (VECMs) are commonly used. To define our set of models, let x_t be an $n \times 1$ vector (i.e., in our case $n = 3$ and $x_t = (c_t, a_t, y_t)'$). A VECM can be written as (see, e.g., Johansen 1995):

$$\Delta x_t = \gamma\alpha' x_{t-1} + d_t\mu + \Gamma(L)\Delta x_{t-1} + \varepsilon_t, \tag{2}$$

where γ and α are $n \times r$ matrices with $0 \leq r \leq n$ being the number of co-integrating relationships.[3] $\Gamma(L)$ 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) defines a set of models that differ in the number of co-integrating vectors (r), lag length (p) and the specification of deterministic terms. With regard to the latter, a deterministic trend in the co-integrating 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 μ into these two different parts as:

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

where $\mu_1 = (\gamma'\gamma)^{-1}\ \gamma'\mu$ and $\mu_1 = (\gamma'_\perp\gamma_\perp)^{-1}\gamma'_\perp\ \mu$ (and $\gamma_\perp$ is orthogonal to γ). With this transformation, μ_1 reflects deterministic terms in the co-integrating residual

3. 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).

TABLE 1
SPECIFICATION FOR DETERMINISTIC TERMS

Code	In co-integrating residual	In Δx_t
$d = 1$	$d_t = (1, t)$	$d_t = (1, t)$
$d = 2$	$d_t = (1, t)$	$d_t = (1)$
$d = 3$	$d_t = (1)$	$d_t = (1)$
$d = 4$	$d_t = (1)$	$d_t = (0)$
$d = 5$	$d_t = (0)$	$d_t = (0)$

while μ_2 reflects those in Δx_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 co-integrating 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 Table 1.[4]

Limiting the number of variables affected by the co-integrating residuals (in econometric language these are referred to as weak exogeneity restrictions) is another issue that is commonly considered in this literature. For instance, in the most important of their results, LL impose weak exogeneity of c_t and y_t in the co-integrating relationship. That is, they set the coefficients in γ 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 γ. 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 by the wealth variable.

We have now defined a set of statistical models (i.e., a model space) that vary in their number of co-integrating vectors (r), lag length (p), deterministic terms (d), and whether or not weak exogeneity is imposed. But how should we link this class of statistical models with the theoretical derivations presented at the beginning of this section? These derivations have strong implications for some aspects of model choice, but not others. In particular, they say that we should have $r = 1$ or 2 and the co-integrating residual should not have a trend in it (i.e., $d > 2$). However, the theoretical derivations say nothing about the deterministic term in Δx_t, nor about lag length nor about weak exogeneity. One can imagine three different reactions to this state of affairs. First, 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

4. See Johansen (1995, pp. 81–84) for a motivation for these particular five choices (which involves eliminating some potentially observationally equivalent combinations of deterministic terms).

a non-informative prior over model space. Second, 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. Third, one could argue that all models should be considered, but that more weight should be attached to models that 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.

At a methodological level, the point we are trying to make is that the Bayesian methodology allows for the formal incorporation of economic theory via priors. That is, the theoretical derivations provide us with prior information, either about the model space (e.g., the model with one co-integrating vector involving no deterministic trend in the co-integrating residual is more plausible *a priori*) or about the parameter space. In this sense, Bayesian methods can be used as a tool that allows us to investigate the relative roles of economic theory, the data and the econometric methods in producing empirical conclusions. In this paper, the "economic theory" is provided by the budget constraint and associated assumptions (e.g., linearization and stationarity assumptions) that are used to derive the previous co-integrating relationship. We will refer to statistical models that are consistent with this theory as "consistent with the log-linearized budget constraint." These are models with $r = 1$ or 2 and $d > 2$.

For the reader who strongly believes in the accounting identities that underlie the log-linearized budget constraint, it is worth noting that there are many reasons why there may be statistical evidence in favor of models that apparently violate them. The theoretical derivations provided above assume key parameters are constant and key variables are stationary. But these assumptions might be bad for many reasons (e.g., if the growth rate increases and the relationship between future and current income is not stable). Hence, models that are not consistent with the log-linearized budget constraint do not necessarily violate basic accounting identities.

Our econometric methods, which are largely the same as those developed in Strachan (2003), Strachan and Inder (2004), and Strachan and van Dijk (2006), 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), to obtain posterior properties of a function of interest (e.g., a variance decomposition), we need $p(\phi \mid Data)$ that depends on $p(M_r \mid Data)$ and $p(\phi \mid Data, M_r)$. Properties of the latter two densities can be obtained from the posterior simulators for each model described in the Appendix. $p(M_r \mid Data)$ and $p(\phi \mid Data, M_r)$ depend on the likelihood function (defined by equation (2), 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. Relating to the prior over parameter space we adopt four different approaches. First, we use a standard non-informative prior to produce empirical results in a specific model, but use the Schwarz criterion (BIC) to approximate $p(M_r \mid Data)$ when averaging across models.[5]

5. The difference between BICs for two models is approximately equal to the log of the Bayes factor comparing the two models. This relationship can be used to approximate $p(M_r \mid Data)$.

Such an approach has the attraction that the researcher does not need to subjectively elicit any prior hyper-parameters. It has been used in many model averaging papers and, in particular, is popular with researchers doing Bayesian averaging of classical estimates (BACE); see Sala-i-Martin, Doppelhoffer, and Miller (2004). However, it is an approximate method and has been criticized by, among others, Berger and Pericchi (1996, 2001).[6] This approach does not suffer from Bartlett's paradox.[7]

Second, we use the fractional Bayes factor approach of O'Hagan (1995). This involves selecting a fraction of the data (a training sample) that combines with a non-informative prior to yield a posterior. This posterior is then used as a "prior" that is combined with the remainder of the data. Since the "prior," is proper, a valid marginal likelihood is obtained. The fractional Bayes factor approach is related to our third approach that uses the instrinsic Bayes factor (see, e.g., Berger and Pericchi 1996). The intrinsic Bayes factor also involves using training samples to create a "prior," which is then combined with the remainder of the data. However, the intrinsic Bayes factor involves averaging over Bayes factors calculated using each possible training sample (see the Appendix for further details).

Fourth, 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.[8] For the prior over the co-integration space, we choose the reference prior described in Strachan and Inder (2004).[9] This prior is uniform over the co-integration space and, thus, non-informative in the sense described in Strachan and Inder (2004). But, the co-integration space is compact and, thus, we do not run into Bartlett's paradox (see Strachan and van Dijk 2003, for further discussion).

In addition, we present results averaged over the four approaches. 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 co-integrating vectors), but also by prior assumptions and it makes sense to present a grand average over all such models.[10] Thus, we are using a wide variety of objective and reference Bayesian

6. For instance, Berger and Pericchi (1996) note that BIC will not typically correspond to a Bayesian analysis in finite samples (but will asymptotically).

7. Bartlett's paradox occurs when improper priors (which depend on arbitrary normalizing constants) are used to compare models or calculate $p(M_r \mid Data)$. Since improper priors are often "non-informative," this is a worry for the objective Bayesian and motivates our use of either approximate methods (i.e., BIC) or methods that use proper priors. None of the priors used in this paper suffer from this problem.

8. In Bayesian jargon, we use a hierarchical prior for the shrinkage parameter.

9. A reference prior is one that 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 co-integration 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 (2005).

10. We use a simple average with equal weights for the BIC, fractional Bayes factor, intrinsic 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.

approaches to model comparison. We argue that this sort of approach is crucial in order to investigate the prior robustness of any empirical results.

It is worthwhile to digress briefly to discuss the interpretation of model averaged results when the true model is misspecified. There are some subtle methodological issues here since some Bayesians would argue that the concept of a true model (and, thus, misspecification) is not a sensible one (i.e., models can be better or worse at explaining the data, but not true or false). Furthermore, Bayesian model averaging can be interpreted as working with a single model, but the single model is a big mixture model involving all the individual models that define the model space. In this sense, the "model" produced by averaging over our 84 VECMs can approximate a much wider range of behaviors than any of the individual VECMs and, thus, is more likely to be "true." However, it is worth noting that, if there is a true model that is not in our set of models, the results of Berk (1966), Dmochowski (1996), and Fernandez-Villaverde and Rubio-Ramirez (2004) imply that use of Bayes factors to select models will, asymptotically, result in the selection of the model that is closest (in Kullback-Leibler distance sense) to the true one. Thus, asymptotically, our approach will set $p(M_r \mid Data) = 1$ if M_r is either the true model or the closest to the true model.

Note also that model averaging can yield results that are the same as model selection in two ways. First, if one model (say M_r) is clearly preferred by the data then we will have $p(M_r \mid Data) \approx 1$ (and alternative models receiving weights near zero). Second, even if many models receive appreciable support from the data, it is possible that all of them will have similar implications for the economic features of interest (e.g., variance decompositions or impulse responses). We discuss these two possibilities below.

2. EMPIRICAL RESULTS

Our empirical results use data on c_t, which is the log of total real personal consumption expenditures; a_t, which is the log of real household net worth (including all financial and household wealth but excluding consumer durables); and y_t, which is the log of after-tax labor income. We use the PCE deflator to obtain real values, which we measure in 2006Q2 dollars. We use the timing convention from Lettau and Ludvigson (2001), that is, current quarter wealth is the value at the end of the previous quarter. Unlike Lettau and Ludvigson, we use total consumption expenditures and remove the value of durables from net worth. This follows the recommendation of Rudd and Whelan (2006) that the log-linearized budget constraint only strictly applies to this definition of the data. Our data run from 1951Q4 through 2006Q2.[11]

11. The working paper version of this paper, available at http://personal.strath.ac.uk/gary.koop/, uses the LL data set and finds results similar to those presented here (although the imposition of weak exogeneity becomes even more important in recovering their results). We also worked with the 1951Q4 through 1995Q4 sub-sample considered by LL. These subsample results are qualitatively similar to those obtained using the full sample.

2.1 Model Comparison Results

In this section, we present results relating to the posterior model probabilities, $p(M_r \mid Data)$. As described in the previous section, models are defined by the number of co-integrating 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 ($d = 1, \ldots, 5$).[12] For comparison, remember that the log-linearized budget constraint implies that there should be one (or more) co-integration relationships and no deterministic trend in the co-integration relationship. LL, using a slightly different data set, chose a model consistent with these considerations (i.e., $r = 1$, $d = 3$ and $p = 1$).[13]

Given the large number of models and four different approaches to model comparison (i.e., based on BIC, fractional and intrinsic 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 co-integrating vectors (integrating over deterministic terms and lag length), lag length (integrating over r and d) and deterministic trends (integrating over r and p) for our different approaches. In this table, we do not attach any extra weight to models consistent with the log-linearized budget constraint, but simply allocate equal prior weight to each model.

With regard to the shrinkage prior, as described in the Appendix, this requires the choice of a shrinkage parameter that we call v. 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 v where the weights in the averaging are data based).

A finding that leaps out is that there is a great deal of model uncertainty. With the exception of the shrinkage prior approach, we are not in the case where $p(M_r \mid Data) \approx 1$ for a single model and, thus, model averaging and model selection can potentially yield very different results.

For most approaches, there exists uncertainty over the number of co-integrating vectors and deterministic trends. For no approach is the single co-integrating vector case the most popular choice. However, just which value for r is preferred varies across approaches. The intrinsic Bayes factor and shrinkage prior approaches provide support for $r = 0$ and, thus, that co-integration does not occur. In contrast, the BIC and fractional Bayes factor approaches indicate that two co-integrating vectors are more probable than one.

12. Note that, in some cases, two different models are observationally equivalent (e.g., if $r = 0$, the treatment of deterministic terms in the co-integrating residual is irrelevant since a co-integrating residual does not exist). We can choose to leave observationally equivalent models in to make sure each value for r receives equal weight, *a priori*. Alternatively, we can choose to count only one of the observationally equivalent models in to make sure each distinct statistical model receives equal weight, *a priori*. Although the latter gives more *a priori* weight to co-integrating models, we adopt this approach. In any case, there was almost no difference in the posterior results between the two approaches.

13. 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 co-integrating vectors and lag length. LL use a different notation for lags so that their "2 lags" refers to 2 lags in the unrestricted VAR. This corresponds to our $p = 1$ for lagged differences in the VECM. LL do not provide a detailed presentation of results relating to choice of deterministic terms.

TABLE 2

Results Using Uniform Prior over Space of All Models

	BIC	Fractional Bayes fact.	Intrisic Bayes fact.	Shrinkage prior	Average over all approaches
			Results for co-integrating rank: $p(r \mid Data)$		
$r = 0$	0.000	0.026	0.962	1.000	0.497
$r = 1$	0.111	0.294	0.037	0.000	0.111
$r = 2$	0.850	0.635	0.000	0.000	0.371
$r = 3$	0.039	0.045	0.000	0.000	0.021
			Results for lag length: $p(p \mid Data)$		
$p = 0$	0.000	0.000	0.954	1.000	0.489
$p = 1$	0.976	0.999	0.046	0.000	0.506
$p = 2$	0.022	0.001	0.000	0.000	0.006
$p = 3$	0.002	0.000	0.000	0.000	0.001
			Results for deterministic trends: $p(d \mid Data)$		
$d = 1$	0.111	0.047	0.013	0.000	0.043
$d = 2$	0.796	0.700	0.456	0.000	0.488
$d = 3$	0.048	0.120	0.456	0.000	0.156
$d = 4$	0.045	0.130	0.019	0.000	0.049
$d = 5$	0.000	0.003	0.056	1.000	0.265

With regard to the issue of deterministic trends, there is also a high degree of uncertainty. Certainly, all of our approaches indicate substantial support for $d = 3, 4,$ or 5 (the choices that rule out a deterministic trend in the co-integrating residual). But, with the exception of the shrinkage prior approach, there is also substantial weight to the (economically non-intuitive) cases where there is a time trend in the co-integrating residual.

Table 3 presents some additional empirical evidence on certain key models and hypotheses using our four approaches to Bayesian model averaging. The row labeled $\Pr(LL \mid Data)$ is the probability of the single model used in LL. The row labeled $\Pr(Theory \mid Data)$ is the total probability associated with all models that are consistent with the log-linearized budget constraint (i.e., $r = 1$ or 2 and $d > 2$). All of the probabilities in either of these rows are below 25% and most are much lower. That is, there is very little evidence in favor of either LL's model or the entire set of models consistent with the log-linearized budget constraint. On the other hand, it is not possible to completely rule out the latter set of models since some of our approaches allocate substantial probability to them.

TABLE 3

Results for Particular Models and Hypotheses Using Uniform Prior over Space of All Models

	BIC	Fractional Bayes fact.	Intrisic Bayes fact.	Shrinkage prior	Average over all approaches
$\Pr(LL \mid Data)$	0.029	0.086	0.000	0.000	0.026
$\Pr(Theory \mid Data)$	0.091	0.229	0.038	0.000	0.089
$\Pr(wex \mid r = 1, Data)$	0.096	0.907	0.872	1.000	0.719

To explain the remaining row in Table 3, a brief digression on weak exogeneity is required. 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 most of their results) is that the coefficients in γ corresponding to the equations for c_t and y_t are zero. In words, the immediate impact of the co-integrating residuals is only on assets and not on consumption or income. This hypothesis only makes sense when co-integration occurs and $r = 1$. When $r = 0$ these coefficients do not enter the model. Furthermore, econometric theory tells us that, with n variables and r co-integrating relationships, there can be at most $n - r$ weak exogeneity restrictions. Thus, when $r = 2$ then at most 1 variable can be weakly exogenous. The weak exogeneity assumption of most interest involves two restrictions and, thus, is only relevant for $r = 1$. Accordingly, the row labeled $\Pr(wex \mid r = 1, Data)$ presents the probability of the weak exogeneity assumption using only models where $r = 1$. With the exception of the BIC approach, we are finding strong support for this weak exogeneity restriction when $r = 1$ (but, of course, Table 2 suggests there is a great deal of uncertainty about whether $r = 1$).

The previous results are all based on a non-informative prior over the model space. That is, each of our 84 models received an *a priori* weight of $\frac{1}{84}$. Given that the theory based on the linearization of the budget constraint suggests that co-integration is present ($r = 1$ or 2) and $d > 2$ (i.e., theory suggests 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 the log-linearized budget constraint should receive twice the weight of other models, then she should attach prior weight of $\frac{2}{120}$ to the 36 models consistent with it and $\frac{1}{120}$ to the 48 models, which are not. Results in Table 2 could be adjusted appropriately.

A third strategy a researcher might take is to impose the log-linearization of the budget constraint 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 4. These results are more consistent with standard approaches used in the literature (e.g., the model of LL). Note, however, that for the BIC and fractional Bayes factor approaches, there still is substantial probability attached to two co-integrating vectors. We will see whether these differences have substantial economic implications in the next section on variance decompositions and impulse response functions.

2.2 Variance Decompositions and Impulse Responses

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

TABLE 4

RESULTS USING UNIFORM PRIOR OVER MODELS CONSISTENT WITH LOG-LINEARIZED BUDGET CONSTRAINT

	BIC Bayes fact.	Fractional Bayes fact.	Intrisic bayes fact.	Shrinkage prior	Average over all approaches
			Results for co-integrating rank: $p(r \mid Data)$		
$r = 1$	0.328	0.394	0.994	1.000	0.679
$r = 2$	0.672	0.606	0.006	0.000	0.321
			Results for lag length: $p(p \mid Data)$		
$p = 0$	0.000	0.000	0.995	1.000	0.499
$p = 1$	0.994	1.000	0.005	0.000	0.500
$p = 2$	0.006	0.000	0.000	0.000	0.002
$p = 3$	0.000	0.000	0.000	0.000	0.000
			Results for deterministic trends: $p(d \mid Data)$		
$d = 3$	0.519	0.446	0.001	0.000	0.242
$d = 4$	0.479	0.543	0.015	0.498	0.384
$d = 5$	0.002	0.010	0.984	0.502	0.375

uncertainty is substantial would be unimportant. Accordingly, in this section we investigate the consequences of model uncertainty for functions of interest to economic policymakers. For the sake of brevity, we focus on variance decompositions and impulse responses. Researchers working in this area have presented a wider battery of results, but typically variance decompositions are at the heart of the story. Briefly, co-integration restrictions allow us to decompose x_t into permanent and transitory components. We can then measure the role each of these plays. We use the permanent-transitory decomposition of Centoni and Cubadda (2003).[14] Given this decomposition, we can calculate the fraction of the total variance of the forecast error at horizon h, which is attributable to the permanent component (and the fraction attributable to transitory shocks is one minus this).[15] 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 when measuring the role of permanent shocks.

The key questions of practical interest are whether permanent shocks have little role to play in driving changes in wealth (which relates to the variance decomposition) and whether permanent shocks are the main driver of consumption. Hence, although we will mention a range of findings, our main focus will be on this issue. Since our variance decompositions are qualitatively similar at all forecast horizons, we only present results for $h = \infty$.

Table 5 summarizes the posterior properties of the variance decomposition (averaging over all models). We present the posterior mean and median (two commonly

14. Based on simulation comparisons between this and the approach of Gonzalo and Ng (2001), we suspect that the two ostensibly different methods produce identical results for the variance decomposition presented later.

15. This statement assumes that the transitory and permanent shocks are orthogonal, a common assumption used by many (including LL).

TABLE 5

Variance Decompositions Using All Models: Share of Forecast Error Due to Permanent Component

	Wealth	Consumption	Income
	BIC		
Mean	0.188	0.378	0.696
Median	0.145	0.335	0.805
		[0.015,0.435]	[0.000,0.000]
50% HPDI	[0.000,0.185]	[0.985,1.000]	[0.815,1.000]
IQR	[0.036,0.295]	[0.116,0.615]	[0.545,0.925]
	Fractional Bayes factor		
Mean	0.272	0.435	0.573
Median	0.225	0.275	0.535
		[0.000,0.115]	[0.275,0.525]
50% HPDI	[0.000,0.245]	[0.955,1.000]	[0.925,1.000]
IQR	[0.056,0.445]	[0.076,0.895]	[0.326,0.885]
	Intrinsic Bayes factor		
Mean	0.991	0.961	0.977
Median	1.000	1.000	1.000
50% HPDI	[1.000,1.000]	[1.000,1.000]	[1.000,1.000]
IQR	[1.000,1.000]	[1.000,1.000]	[1.000,1.000]
	Shrinkage prior		
Mean	1.000	1.000	1.000
Median	1.000	1.000	1.000
50% HPDI	[1.000,1.000]	[1.000,1.000]	[1.000,1.000]
IQR	[1.000,1.000]	[1.000,1.000]	[1.000,1.000]
	Average over all approaches		
Mean	0.611	0.692	0.810
Median	1.000	1.000	1.000
50% HPDI	[1.000,1.000]	[1.000,1.000]	[1.000,1.000]
IQR	[0.175,1.000]	[0.295,1.000]	[0.675,1.000]

used point estimates) as well as two measures of the dispersion of the posterior. These are a 50% Highest Posterior Density Interval (HPDI) that is defined as the shortest interval containing 50% of the posterior probability and an inter-quartile range (IQR): the 25th and 75th percentiles of the posterior.[16] 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 inter-quartile range should be as well. Clearly they are not in our case. In fact, the HPDIs are sometimes discontinuous, made up of two disjoint intervals. This property is due to the fact that BMA involves averaging different distributions—the result can be a multi-modal 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

16. We calculate these by creating a histogram defined over a grid and then using its properties to derive HPDIs and IQRs.

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, with consumption and income (using the BIC approach), the 50% HPDI intervals contain 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 these variables or they play the dominant role.

A key conclusion from Table 5 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 consumption. Our different approaches are yielding very different results, 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. As an example, consider the variance decomposition for the wealth variable produced using BIC. It is the one that yields results that are closest to the "transitory shocks are playing the predominant role in driving wealth" story. But the upper bound of the IQR is approximately 0.3, indicating there is a 25% chance that permanent shocks are playing a quite substantial role in driving wealth. The approaches that find little evidence for co-integration (i.e., the intrinsic Bayes factor and shrinkage prior) say that the permanent shocks are (almost exclusively) the ones driving fluctuations in wealth.

This finding can be seen more clearly in Figure 1 that, for the wealth variable, plots the posterior distributions of the variance decomposition (i.e., it plots the full

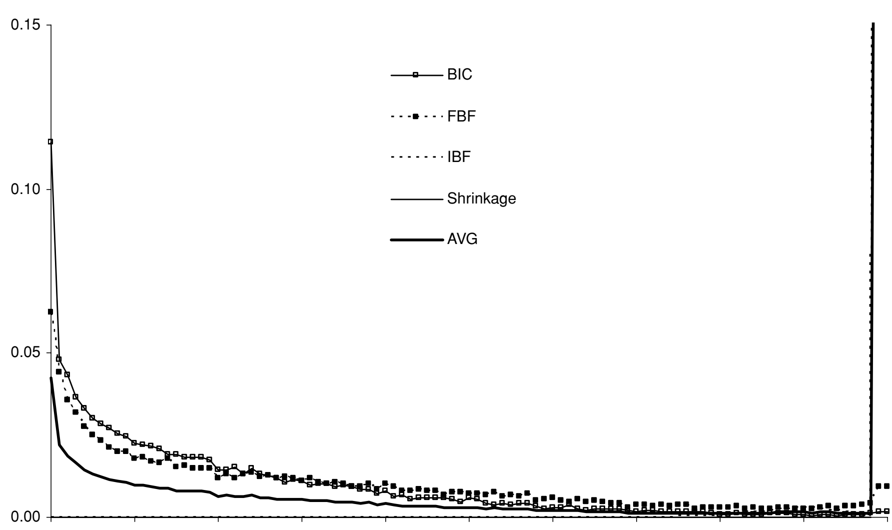

FIG. 1. Plots of the Densities of the Proportion of Wealth Variability Due to Permanent Shocks, Averaged Over All Models (i.e., all r, all d, and all p).

NOTE: There is a spike at one for each density. The heights at one of each density are: 0.0015 for BIC, 0.0090 for FBF, 0.6838 for IBF, 0.9986 for the shrinkage prior, and 0.4378 for the averaged density.

TABLE 6

VARIANCE DECOMPOSITIONS MODELS CONSISTENT WITH LOG-LINEARIZED BUDGET CONSTRAINT: SHARE OF FORECAST ERROR DUE TO PERMANENT COMPONENT

	Wealth	Consumption	Income
	BIC		
Mean	0.261	0.453	0.552
Median	0.215	0.365	0.525
	[0.000,0.195]	[0.000,0.125]	[0.255,0.545]
50% HPDI	[0.225,0.245]	[0.945,1.000]	[0.915,1.000]
	Fractional Bayes factor		
Mean	0.276	0.433	0.573
Median	0.225	0.265	0.545
	[0.000,0.195]	[0.000,0.115]	[0.275,0.545]
50% HPDI	[0.235,0.275]	[0.955,1.000]	[0.915,1.000]
	Intrinsic Bayes factor		
Mean	0.886	0.091	0.531
Median	0.896	0.086	0.536
50% HPDI	[0.865,0.945]	[0.035,0.105]	[0.485,0.595]
	Shrinkage prior		
Mean	0.889	0.089	0.537
Median	0.896	0.086	0.536
50% HPDI	[0.875,0.945]	[0.035,0.105]	[0.495,0.605]
	Average over all approaches		
Mean	0.578	0.267	0.548
Median	0.815	0.116	0.536
	[0.000,0.015]	[0.000,0.105]	[0.425,0.625]
50% HPDI	[0.835,0.965]	[1.000,1.000]	[1.000,1.000]

posteriors which were used to create Table 5). All of the posteriors have a some mass in LL's region, but are fairly flat. Furthermore, with the exception of BIC and the fractional Bayes factor, there is an appreciable mode at 1 (i.e., the point where the permanent component is completely dominant—the point exactly opposite of the "permanent shocks have little role to play in driving wealth" story). Thus, although there is some support for the view that transitory shocks dominate fluctuations in wealth, this support is quite weak when we simply average over all of our models.

For the consumption and income variables, the approaches that indicate strong support for $r = 0$ will, by definition, indicate strong support for the view that permanent shocks that are playing the predominant role. But, similar to the case of the wealth variable, the other approaches have posteriors that exhibit substantial dispersion and multi-modality

Table 6 presents the same information as Table 5, except that we average only over models consistent with the log-linearized budget constraint. 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 alters the posterior

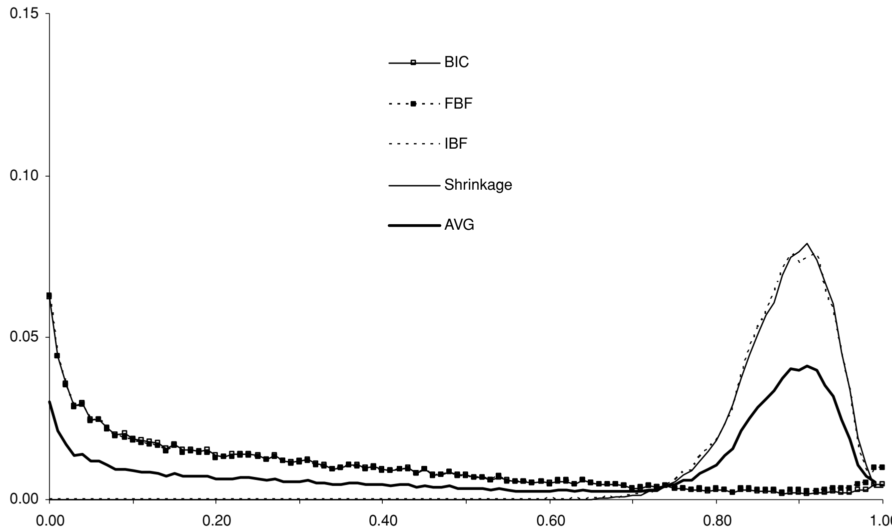

FIG. 2. Plots of the Densities of the Proportion of Wealth Variability Due to Permanent Shocks, Averaged over All Models Consistent with the Log-Linear Budget Constraint (i.e., $r = 1, 2, d > 2$, and all p).

distributions of the variance decomposition of wealth in a direction that is slightly more consistent with the story "permanent shocks have little role in driving wealth." Most importantly, by ruling out $r = 0$, we are making the spikes in the posteriors at the point 1 much smaller. However, these spikes at one are being replaced by appreciable posterior probabilities in the interval [0.8,1.0] for wealth. Note, too, that imposing the log-linearized budget constraint also moves the posterior distribution of the variance decomposition of consumption toward a greater role for transitory shocks.

To summarize the story so far: when we work with the shrinkage prior and intrinsic Bayes factor approaches we are finding very little support indeed for the view that transitory shocks are driving wealth. Using the BIC or fractional Bayes factor approaches we are finding more support for this view, but even in these cases the variance decompositions are imprecisely estimated and the variance decomposition for consumption suggests transitory shocks are important. It is also important to note that these conclusions hold regardless of whether we work with all models or only those consistent with the log-linearized budget constraint.

Thus far we have not considered the weak exogeneity restriction that transitory shocks have no immediate effect on c_t and y_t. For the reasons give above, this restriction can only be imposed on models with $r = 1$. Before discussing this restriction, we note that when we work only with models that have $r = 1$ and do variance decompositions as above, their posteriors look very similar to those in Figure 2. For the sake of brevity, we do not plot such figures, but stress only that imposing one co-integrating vector does not alter our empirical results substantively (relative to the case where we work with all models consistent with the log-linearized budget constraint).

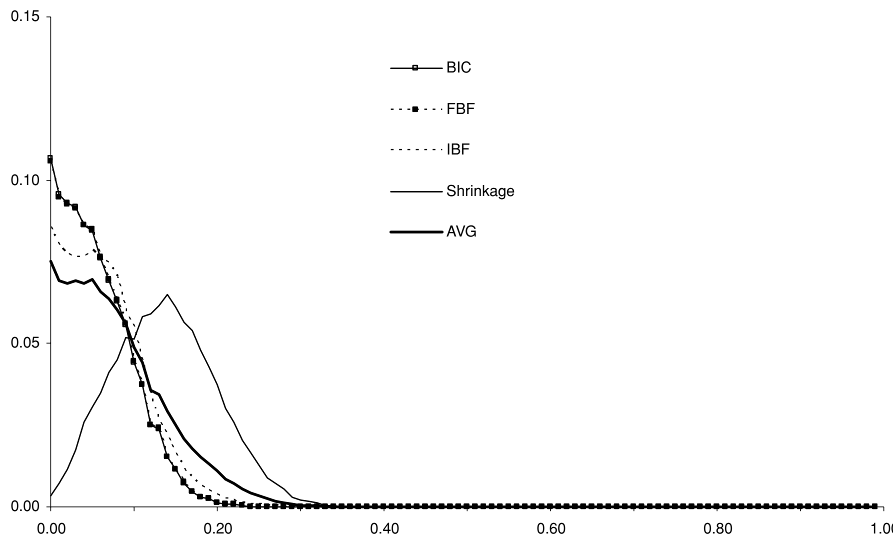

FIG. 3. Plots of the Densities of the Proportion of Wealth Variability Due to Permanent Shocks, Averaged over Models Consistent with Log-Linear Budget Constraint (with $r = 1$, $d > 2$, and all p) *with* Weak Exogeneity.

Figure 3 is in the same format as Figures 1 and 2. However, it averages over models that are consistent with the log-linearized budget constraint, have $r = 1$ and have the weak exogeneity restrictions imposed. In this case we are effectively forcing consumption to be mainly driven by permanent shocks. Note that now, for all of our approaches, we are finding evidence in support of the story "permanent shocks have little role in driving wealth and, thus, transitory shocks predominate." Thus, we are finding that this weak exogeneity restriction is the crucial factor in recovering this story. We remind the reader of the findings in Table 3 that, *if one co-integrating vector is imposed*, then this weak exogeneity restriction does receive support. However, from Table 2, it is clear that there is not strong support for $r = 1$.

The evidence we have provided so far suggests that there is considerable uncertainty as to exactly how much of the variation in consumption and wealth is driven by permanent shocks. This is only the first part of the story. We now consider the implications of this uncertainty for the behavior of consumption by presenting generalized impulse response functions of the response in consumption to shocks in wealth. We consider two sorts of shocks. One is a conventional 10% positive shock to wealth. The other is the transitory component of this shock. Thus, we can measure the total effect of a shock to wealth on consumption and the effect only of the transitory part of that shock. We identify the impulse responses using an ordering convention. Note that our data is timed such that assets are recorded on the last day of the quarter so that, e.g., wealth measured at the end of 2006Q1 is matched with 2006Q2 consumption. This helps motivate our ordering convention of a_t, c_t, y_t. Thus, a shock to consumption or income at time t does not effect assets at time t.Our figures present posteriors of

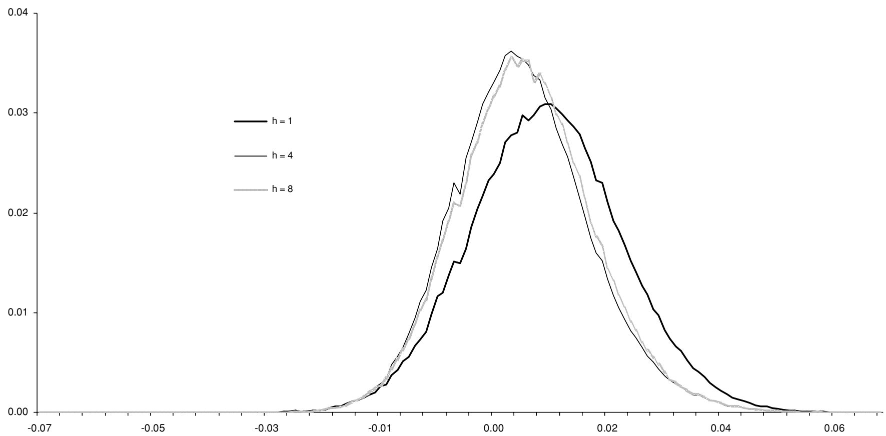

FIG. 4. Average Plots of the GIRF for the Response in Consumption to a 10% Shock to Wealth.

NOTE: The densities are averaged over all models (i.e., all r, all d, and all p).

impulse responses at horizons of $h = 1$, 4, and 8. The generalized impulse response function is the answer to the question, "If wealth just increased by 10% how does that change the forecast for consumption?"

As we did for the variance decompositions, we carry out three Bayesian model averaging exercise. The first averages over all models. The second averages over all models consistent with the log-linearized budget constraint. The third averages over all models with all models consistent with the log-linearized budget constraint with the additional restrictions that $r = 1$ and that weak exogeneity is imposed. For the sake of brevity, we do not present results for each of our four approaches (i.e., based on BIC, fractional and intrinsic Bayes factors and the shrinkage prior) but rather present an equally weighted average over all of these.

Let us first consider the total effect of a shock to wealth on consumption for our three BMA exercises. Posteriors for the relevant impulse responses are plotted in Figures 4, 6, and 8. These figures are all similar to one another. The posterior modes in all figures, for all of $h = 1$, 4, and 8, are in the region of 0.01. This value suggest that there will be a 1% increase in consumption in response to a 10% shock in wealth. Using data from quarter 2 of 2006, this implies that a \$1 increase in wealth is likely to produce approximately a \$0.02 increase in consumption. This is a bit below most point estimates in the literature but not seriously at odds with them. However, all of these posteriors are quite dispersed, indicating that it is hard to pin down the magnitude of the impulse response precisely. For instance, the value of zero (indicating wealth shocks have no effect on consumption) is not an implausible one in any of the figures. On the other hand, the value of 0.04 is also not implausible. This latter value (using data from quarter 2 of 2006) would imply that a \$1 increase in wealth produces over a \$0.06 increase in consumption.

If we now examine the impulse responses to a transitory shock in wealth (Figures 5, 7, and 9), we can see the role that the weak exogeneity has to play. Figures 5 and 7

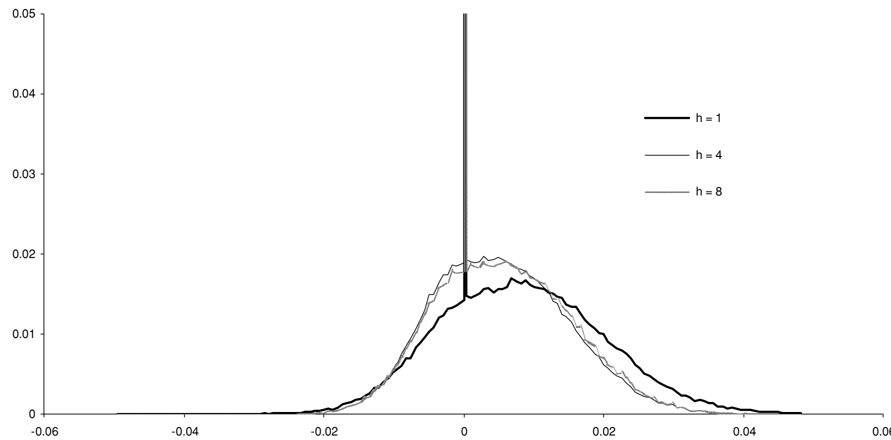

FIG. 5. Average Plots of the GIRF for the Response in Consumption to the *Transitory Component* of a 10% Shock to Wealth.

NOTE: The densities are averaged over all models (i.e., all r, all d, and all p). There is a peak of 0.25 in each density at zero on the y-axis.

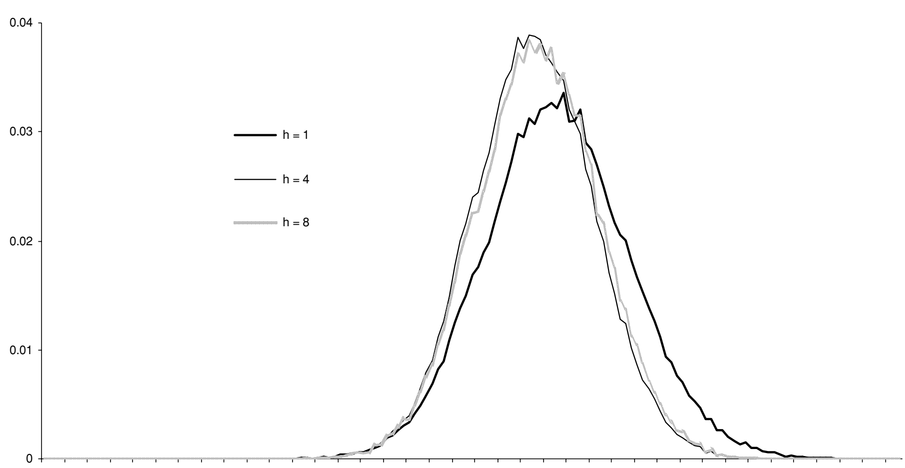

FIG. 6. Average Plots of the GIRF for the Response in Consumption to a 10% Shock to Wealth.

NOTE: The densities are averaged over all models consistent with the log-linear budget constraint (i.e., $r = 1, 2$, $d > 2$, and all p).

(which do not impose weak exogeneity), exhibit a high degree of uncertainty over magnitude of the impulse response.[17] However, when we impose weak exogeneity

17. The spikes in Figure 5 at zero is associated with models with $r = 0$. For these all shocks are permanent so, by definition, transitory shocks play no role.

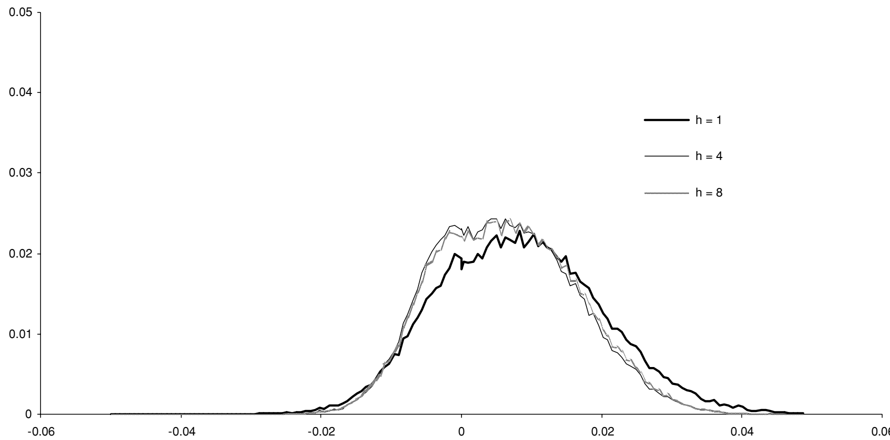

FIG. 7. Average Plots of the GIRF for the Response in Consumption to the *Transitory Component* of a 10% Shock to Wealth.

NOTE: The densities are averaged over all models consistent with the log-linear budget constraint (i.e., $r = 1, 2$, $d > 2$, and all p).

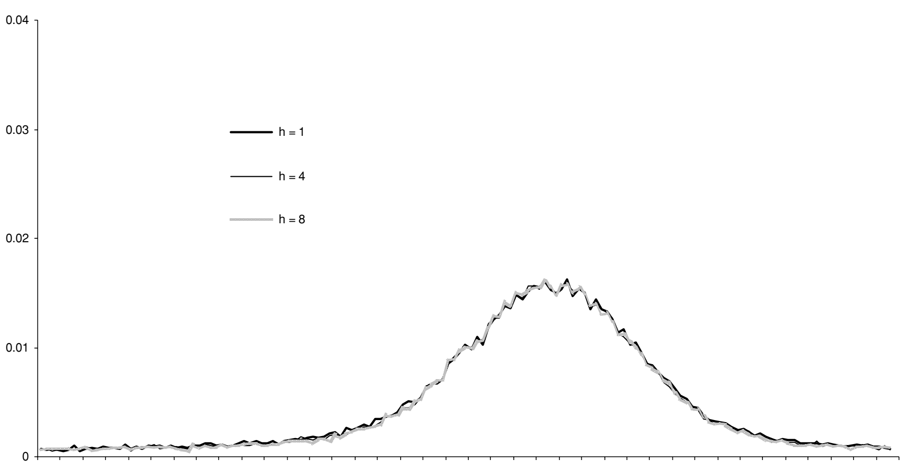

FIG. 8. Average Plots of the GIRF for the Response in Consumption to a 10% Shock to Wealth.

NOTE: The densities are averaged over models consistent with log-linear budget constraint (with $r = 1$, $d > 2$, and all p) *with* weak exogeneity.

(Figure 9), the posterior is much tighter around zero (i.e., the point that says transitory shocks have no effect on consumption). Thus, it is only when we impose this weak exogeneity restriction that we are finding that transitory shocks to wealth do not effect consumption.

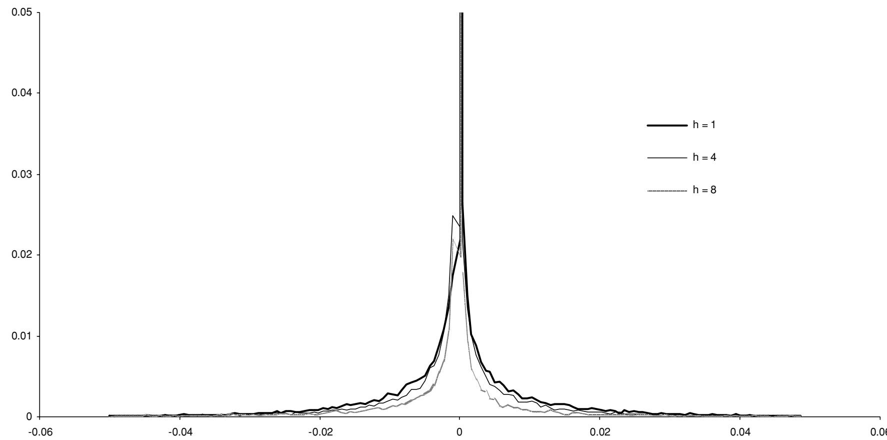

FIG. 9. Average Plots of the GIRF for the Response in Consumption to the *Transitory Component* of a 10% Shock to Wealth.

NOTE: The densities are averaged over models consistent with log-linear budget constraint (with $r = 1$, $d > 2$, and all p) *with* weak exogeneity. There is a peak in the densities of 0.72 for $h = 1$, 0.75 for $h = 4$, and 0.80 for $h = 8$ at zero on the y-axis.

3. CONCLUSIONS

There are many other features that an empirical researcher may be interested in (e.g., other sorts of variance decompositions, long-horizon regressions).[18] However, for the sake of brevity, we end our empirical results here, having established our basic points. These are that model uncertainty is an important issue in this data set and that empirical conclusions depend on the way this issue is treated. To be precise, an attractive story in the literature is that consumption mainly responds to permanent shocks and does not respond to transitory changes in wealth. Further, most of the fluctuations in wealth are transitory. Thus, wealth effects on consumption are small. This story can explain many apparently puzzling recent experiences (e.g., why recent stock market falls have not caused larger changes in consumption). In this paper, we have found that if we simply average over a reasonable set of models, there is great uncertainty over both components of this story. Variance decompositions indicate that it is uncertain whether most of the fluctuations in wealth really are transitory and whether most of the fluctuations in consumption are permanent. Furthermore, impulse response analysis suggests it is unclear that transitory changes in wealth have no effect on consumption. This conclusion does not change when we work only with models consistent with the log-linearized budget constraint suggested by theoretical considerations. However, when we further restrict our models to impose exactly one co-integrating relationship and impose a key weak exogeneity restriction, we do find strong evidence in support of the story.

18. 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).

APPENDIX: BAYESIAN ECONOMETRIC METHODS

The posterior probabilities for each model, $p(M_r \mid Data)$, 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

$$p(M_r \mid Data) = \frac{m_r(y)}{\Sigma_i m_i(y)}. \tag{A1}$$

We have summed over all models in the denominator in (A1). In this paper, we estimate the marginal likelihoods using four methods. First, we adopt the fractional Bayes factor approach of O'Hagan (1995). Second, we use an approach based on the intrinsic Bayes factor of Berger and Pericchi (1996). Third, we use 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 that 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 that is then used as a prior for the remainder of the sample. One feature of this method is the choice of particular sub-sample 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 $b = 17/213$. 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

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

where $\hat{l}_r$ 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 (A1).

The intrinsic Bayes factor approach also involves training samples of size i. As there is some ambiguity as to the meaning of non-contiguous samples in time series models, we took every possible contiguous training sample (e.g., the first training sample uses data $t = 1, \ldots, i$; the second uses data from $t = 2, \ldots, i+1$; through the final training sample from $t = T - i, \ldots, T$) and produced $T - i$ Bayes factors with $i = 100$. We then took the geometric average of these Bayes factors to product the intrinsic Bayes factors. These were then used to construct the posterior model probabilities. The intrinsic Bayes factor approach requires the marginal likelihoods using the training sample and included sample (see Berger and Pericchi, 2001, equation (5)). Closed forms do not exist for these marginal likelihoods and simulation methods would be very time consuming. Therefore these are approximated using BIC.

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

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

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 $p(M_r \mid Data)$ from this expression via (A1). These techniques, which are based upon asymptotic arguments, have the advantages of consistency and simplicity. Furthermore, subjective elicitation of prior hyper-parameters is not required.

The fourth method we use is an approximation to $m(y)$ based upon analytical integration in some dimensions and an 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 (Σ, θ, α) where θ is a vector containing the elements in γ, μ, and $\Gamma(L)$ and Σ is the covariance matrix for the multivariate Normal errors, ε_t. Next, an improper diffuse prior is specified for Σ, a relatively diffuse but proper prior is specified for θ and a flat (proper) prior is specified for α. 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, $p(\Sigma, \theta, \alpha \mid M_r, Data)$, such that

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

As a closed form expression (conditional upon α) exists for the outer integral with respect to (Σ, θ), we can perform this integral analytically. It is the inner integral with respect to α 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 (co-integrating) space of α and α is specified as a semi-orthogonal 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 α . Further details on the computation of $m_r(y)$ by MCMC are provided below and in the survey paper Koop et al. (2006).

A.1 The Prior

We use the following commonly used non-informative prior for Σ

$$p(\Sigma) \propto |\Sigma|^{-(n+1)/2}. \tag{A2}$$

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, Ley, and Steel point out, the prior (A2) is invariant to scale transformations and, although it is not strictly Jeffreys' prior, it is that part of Jeffreys' prior related to Σ and widely accepted as a non-informative prior for Σ.

We next specify a Normal prior for θ as $N(0, \Sigma \otimes I\nu)$ where ν is a hyper-parameter in the variance that controls how diffuse is the prior. It is this hyper-parameter ν that motivates our terminology "shrinkage" prior. As we did not wish to select a particular value for ν but wanted to keep the prior for θ diffuse, we gave ν a Gamma distribution $g(n_1, n_2)$ in which moderately diffuse values for n_1 and n_2 are chosen ($n_1 = 5$, $n_2 = 5$). This prior places 99% of the prior mass for ν in the interval (0.2, 8.4), ensuring the prior for θ will cover a wide range of plausible values.

The prior for α is implied by a uniform prior over the co-integration space. As is frequently stated in empirical co-integration studies, the elements of α 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 co-integrating space, or the space spanned by α. As argued in Villani (2005) and Strachan and Inder (2004), this is not a limitation since it is the co-integrating space that is the object of interest and estimates of α simply provide an estimate of this space. We wish to be non-informative about the co-integrating space and so follow the approach in Strachan and Inder (2004) by specifying α to be semi-orthogonal such that $\alpha' \alpha = I_r$ so α has a compact support, and place a Uniform prior on this support. Alternative specifications may be used to express ignorance about the co-integrating 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.

A.2 Posterior Computation

The prior specification above is also used in Strachan and van Dijk (2006) for a range of model averaging exercises. This specification permits an MCMC sampling scheme called the Gibbs sampler that we describe briefly below.

The Gibbs sampler involves drawing each parameter (or block of parameters) from its posterior distribution conditional upon all the other parameters in the model. Conditional upon α, the model is linear and the posterior for (Σ, θ) has the usual inverted Wishart-Normal form. However, sampling α is complicated by the fact that it is semi-orthogonal. Strachan and van Dijk (2006) adapt the sampler developed in Koop, León-González, and Strachan (2006) to simplify the sampler. The trick is to introduce an unidentified parameter, D, with a proper prior and transform by $\gamma\alpha = \gamma D D^{-1}\alpha = \bar{\gamma}\bar{\alpha}$ such that $\bar{\alpha}$ and $\bar{\gamma}$ have the appropriately dimensioned real space as their support. Next let $\bar{\theta}$ be the vector θ with γ replaced by $\bar{\gamma}$. The conditional posterior distributions of $\bar{\theta}$ and $\bar{\alpha}$ are now Normal and the sampling scheme proceeds as follows.[19]

19. Exact details on the derivations of the parameters of the conditional posteriors can be found in Strachan and van Dijk (2006).

(i) Initialize $(\Sigma, \bar{\theta}, \bar{\alpha}) = (\Sigma^{(0)}, \bar{\theta}^{(0)}, \bar{\alpha}^{(0)})$.
(ii) Draw $\Sigma \mid \bar{\theta}, \bar{\alpha}$ from inverted Wishart.
(iii) Draw $\bar{\theta} \mid \Sigma, \bar{\alpha}$ from the Normal.
(iv) Draw $\bar{\alpha} \mid \Sigma, \bar{\theta}$ from the Normal.
(v) Repeat steps 2 to 4 for a suitable number of replications.

To compute the posterior probabilities of the models we use the Savage–Dickey density ratio (see, e.g., Verdinelli and Wasserman 1995) approach described for this type of models in Koop, León-González, and Strachan (2006). To do this we need a 'base' model, M_0, that is nested within each of the other models in the model set. Note that this model does not have to be actually in the model set. For our case, we consider the model with $d = 5$, $r = 0$, and $p = 0$ as our base model. This model nests within each model M_i at the point where $\bar{\theta} = 0$ and the Bayes factor for the base model M_0 to the model M_i can be computed as

$$B_{0,i} = \frac{p\left(\bar{\theta} \mid M_i, y\right)|_{\bar{\theta}=0}}{p\left(\bar{\theta} \mid M_i\right)|_{\bar{\theta}=0}}.$$

We can use the output from the Gibbs sampler and the conditional posterior $p(\bar{\theta}|\Sigma, \bar{\alpha}, M_i, y)$ and the conditional prior $p(\bar{\theta}|\Sigma, \bar{\alpha}, M_i, y)$ to estimate the required ratio as:

$$\hat{B}_{0,i} = \frac{\frac{1}{M}\Sigma_{m=1}^{M} p\left(\bar{\theta}|M_r, \Sigma^{(m)}, \bar{\alpha}^{(m)}, y\right)|_{\bar{\theta}=0}}{\frac{1}{M}\Sigma_{m=1}^{M} p\left(\bar{\theta}|M_r, \Sigma^{(m)}, \bar{\alpha}^{(m)}\right)|_{\bar{\theta}=0}},$$

where $m = 1, \ldots, M$ denote the (post-burn-in) Gibbs sampler replications and (m) superscripts denote the replications themselves.

As the ratio of posterior probabilities of model M_0 to model M_i is given by

$$\frac{p(M_0 \mid Data)}{p(M_i \mid Data)} = B_{0,i} \frac{p(M_0)}{p(M_i)},$$

where $\frac{p(M_0)}{p(M_i)}$ is the ratio of prior model probabilities, then once we have computed $\hat{B}_{0,i}$ for each model, we can obtain the posterior probabilities for all models of interest.

LITERATURE CITED

Berger, James, and Luis Pericchi. (1996) "The Intrinsic Bayes Factor for Model Selection and Prediction." *Journal of the American Statistical Association*, 91, 109–22.

Berger, James, and Luis Pericchi. (2001) "Objective Bayesian Methods for Model Selection: Introduction and Comparison." In *Model Selection*, edited by Patrick Lahiri, Vol. 38. Institute of Mathematical Statistics Lecture Notes.

Berk, Robert. (1966) "Limiting Behavior of Posterior Distributions when the Model is Incorrect." *Annals of Mathematical Statistics*, 37, 51–8.

Brennan, Michael, and Yihong Xia. (2005) "Tay's as Good as Cay." *Finance Research Letters*, 2, 1–14.

Centoni, Marco, and Gianluca Cubadda. (2003) "Measuring the Business Cycle Effects of Permanent and Transitory Shocks in Cointegrated Time Series." *Economics Letters*, 80, 45–51.

Chikuse, Yasuko. (1990) "The Matrix Angular Central Gaussian Distribution." *Journal of Multivariate Analysis*, 33, 265–74.

Cogley, Timothy, and Thomas Sargent. (2005) "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII US." *Review of Economic Dynamics*, 8, 262–302.

Doan, Thomas, Robert Litterman, and Christopher Sims. (1984) "Forecasting and Conditional Projection using Realistic Prior Distributions." *Econometric Reviews*, 3, 1–100.

Dmochowski, J. (1996) "Intrinsic Priors Via Kullback-Leibler Geometry." In *Bayesian Statistics 5*, edited by Jose Bernardo, pp. 543–549. London: Oxford University Press.

Draper, David. (1995) "Assessment and Propagation of Model Uncertainty (with discussion)." *Journal of the Royal Statistical Society Series B*, 56, 45–98.

Fernández, Carmen, Eduardo Ley, and Mark Steel. (2001) "Model Uncertainty in Cross-Country Growth Regressions." *Journal of Applied Econometrics*, 16, 563–76.

Fernandez-Villaverde, Jesus, and Juan Francisco Rubio-Ramirez. (2004) "Comparing Dynamic Equilibrium Economies to Data: A Bayesian Approach." *Journal of Econometrics*, 123, 153–87.

Garratt, Anthony, Kevin Lee, M. Hashem Pesaran, and Yongcheol Shin. (2003) "Forecast Uncertainties in Macroeconometric Modelling: An Application to the UK Economy." *Journal of the American Statistical Association*, 98, 829–38.

Gelfand, Alan E., and Dipak K. Dey. (1994) "Bayesian Model Choice: Asymptotics and Exact Calculations." *Journal of the Royal Statistical Society, Series B*, 56, 501–4.

Gonzalo, Jesus, and Serena Ng. (2001) "A Systematic Framework for Analyzing the Dynamic Effects of Permanent and Transitory Shocks." *Journal of Economic Dynamics and Control*, 25, 1527–46.

Hjort, Nils, and Gerda Claeskens. (2003) "Frequentist Model Average Estimators." *Journal of the American Statistical Association*, 98, 879–99.

Hodges, James. (1987) "Uncertainty, Policy Analysis and Statistics." *Statistical Science*, 2, 259–91.

Johansen, Soren. (1995) *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

Julliard, Christian. (2005) *Labor Income and Wealth, Department of Economics*. Manuscript, Princeton University.

Kass, Robert, and Adrian Raftery. (1995) "Bayes factors." *Journal of the American Statistical Association*, 90, 773–95.

Kleibergen, Frank, and Richard Paap. (2002). "Prior, Posteriors and Bayes Factors for a Bayesian Analysis of Cointegration." *Journal of Econometrics*, 111, 223–49.

Koop, Gary, Roberto León-González, and Rodney Strachan. (2006) "Bayesian Inference in a Cointegrating Panel Data Model." Department of Economics Working Paper 06/02, University of Leicester.

Koop, Gary, Rodney Strachan, Herman van Dijk, and Mattias Villani. (2006) "Bayesian Approaches to Cointegration." In *Palgrave Handbook of Econometrics: Volume 1 Econometric Theory*, edited by T.C. Mills and K. Patterson. Basingstoke: Palgrave Macmillan.

Leamer, Edward. (1978) *Specification Searches*. New York: Wiley.

Lee, Hangyong. (2002) "Deterministic Trend and Biases in the Estimated Consumption-Wealth Ratio." Working Paper Available at http://www.columbia.edu/∼hl306/cay.pdf.

Lettau, Martin, and Sidney Ludvigson. (2001) "Consumption, Aggregate Wealth and Expected Stock Returns." *Journal of Finance*, 56, 815–49.

Lettau, Martin, and Sidney Ludvigson. (2004) "Understanding Trend and Cycle in Asset Values: Reevaluating the Wealth Effect on Consumption." *American Economic Review*, 94, 276–99.

Min, Chung-ki, and Arnold Zellner. (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, Anthony. (1995) "Fractional Bayes Factors for Model Comparison." *Journal of the Royal Statistical Society, Series B*, 57, 99–138.

Palumbo, Michael, Jeremy Rudd, and Karl Whelan. (2006) "On the Relationships Between Real Consumption, Income and Wealth." *Journal of Business and Economic Statistics*, 24, 1–11.

Poirier, Dale. (1995) *Intermediate Statistics and Econometrics: A Comparative Approach*. Cambridge: The MIT Press.

Potter, Simon. (2001) "Using *A Priori* Information in the Analysis of Cointegrated Systems." Working Paper.

Raftery, Adrian, David Madigan, and Jennifer Hoeting. (1997) "Bayesian Model Averaging for Linear Regression Models." *Journal of the American Statistical Association*, 92, 179–191.

Rudd, Jeremy, and Karl Whelan. (2006). "Empirical Proxies for the Consumption-Wealth Ratio." *Review of Economic Dynamics*, 9, 34–51.

Sala-i-Martin, Xavier. (1997) "I Just Ran Two Million Regressions." *American Economic Review*, 87, 178–83.

Sala-i-Martin, Xavier, Gernot Doppelhoffer, and Ronald Miller. (2004) "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach." *American Economic Review*, 94, 813–835.

Schwarz, Gideon. (1978) "Estimating the Dimension of a Model." *Annals of Statistics*, 6, 461–64.

Strachan, Rodney W. (2003) "Valid Bayesian Estimation of the Cointegrating Error Correction Model." *Journal of Business and Economic Statistics*, 21, 185–95.

Strachan, Rodney W., and Brett Inder. (2004) "Bayesian Analysis of the Error Correction Model." *Journal of Econometrics*, 123, 307–25.

Strachan, Rodney W., and Herman K. van Dijk. (2003) "Bayesian Model Selection with an Uninformative Prior." *Oxford Bulletin of Economics and Statistics*, 65, 863–76.

Strachan, Rodney W., and Herman K. van Dijk. (2006) "Model Uncertainty and Bayesian Model Averaging in Vector Autoregressive Processes." Econometric Institute Report EI 2006-08, Erasmus University Rotterdam.

Villani, Mattias. (2005) "Bayes Reference Analysis of Cointegration." *Econometric Theory*, 21, 326–57.