

The Econometric Consequences of the Ceteris Paribus Condition in Economic Theory*

Herman J. Bierens
Pennsylvania State University & Tilburg University

Norman R. Swanson
Pennsylvania State University

September 1998

Abstract

The *ceteris paribus* condition in economic theory assumes that the world outside the environment described by the theoretical model does not change, so that it has no impact on the economic phenomena under review. In this paper, we examine the econometric consequences of the *ceteris paribus* assumption by introducing a "state of the world" variable into a well specified stochastic economic theory, and we show that the difference between the conditional distribution implied by the theoretical model and the actual conditional distribution of the data is due to different ways of conditioning on the state of the world. We allow the "state of the world" variable to be, alternatively and equivalently, an index variable representing omitted variables, or a discrete random parameter representing a sequence of models. We construct a probability that can be interpreted as the upperbound of the probability that the *ceteris paribus* condition is correct. The estimated upperbound can in turn be interpreted as a measure of the information about the data-generating process that is provided by a theoretical model which is constrained by a set of *ceteris paribus* assumptions. In order to illustrate our findings from both a theoretical and an empirical perspective, we examine a linearized version of the real business cycle model proposed by King, Plosser, and Rebello (1988b).

JEL Classification: C32, C51, C52, E32.

Keywords: Ceteris paribus; missing variables; Bayesian prior; information measure; reality bound; fit; stochastic general equilibrium models; real business cycle models.

1 Introduction

In this paper we examine some econometric implications of the *ceteris paribus* assumption (other things being equal, or all else remaining the same) which is used in the construction of economic theories. Most economic theories, including general equilibrium theory, are "partial" theories in the sense that only a few related economic phenomena are studied. For example, supply and demand theories, and theories of the behavior of economic agents within the context of stylized economies that have no direct counterpart in reality are commonplace in economics. In particular, consider the stylized Robinson Crusoe type economy, where a single rational economic agent decides how much of his crop of potatoes he should eat, how much he should plant for next year's harvest, and how long he should work in the fields, in order to maximize his life-time utility. The analysis of this "partial" theory is justified, explicitly or implicitly, by the *ceteris paribus* assumption. However, when simple economic models of this type are estimated using data which are themselves generated from a much more complex real economy, it is not surprising that they often fit poorly. In such situations, some theorists blame the messenger of the bad news, econometrics, for the lack of fit, and abandon econometrics altogether, in favor of simple calibration techniques. (See Hansen and Heckman (1996) for a review of calibration, and Sims (1996) and Kydland and Prescott (1996) for opposite views on calibration.) In this scenario, one may well ask what the role of econometrics is, and why calibration has become so popular. In the following, we attempt to shed light on this issue by formalizing the *ceteris paribus* assumption in economic theory and econometrics. Along these lines, one of our main goals is to provide new evidence concerning the link between economic theory and econometrics.

We first relax the *ceteris paribus* assumption and introduce a "state of the world" variable W . Given a vector Y of dependent variables, and a vector X of predetermined variables, let the conditional density of Y , given $X = x$ and $W = w$, be $f(y|x, w)$. Now impose the *ceteris paribus* assumption in a theoretical economic model by conditioning on the event that the "state of the world" variable W is constant, say $W = 0$, which yields the theoretical conditional density $f_0(y|x) = f(y|x, 0)$. On the other hand, the true conditional density of Y , given $X = x$ alone is $f(y|x) = E[f(y|x, W)]$. Therefore, the *ceteris paribus* condition will in general cause misspecification of the theoretical model $f_0(y|x)$. This is the very reason

why theoretical models often don't fit the data.

Next, we allow the "state of the world" variable W to be either: (1) a vector of omitted variables, which may be represented, without too much loss of generality, by a discrete scalar random variable, or (2) a discrete random parameter, with a "prior" which is the maximum probability that the ceteris paribus condition holds. We show that both interpretations of the "state of the world" lead to the same link between the theoretical model and the data-generating process, and are therefore observationally equivalent. In particular, we show that under some regularity conditions one can write $f(y) = p_0 f_0(y|\beta_0) + (1 - p_0) f_1(y|\beta_0)$, where $f(y)$ is the (conditional) density of the data-generating process, $f_0(y|\beta)$ is a parametrization of the (conditional) density $f_0(y)$ implied by the theoretical model, with β a parameter vector, and p_0 is the maximum number between zero and one for which $f_1(y|\beta_0)$ is a density¹. In other words, $p_0 = \sup_{\beta} \inf_y f(y)/f_0(y|\beta)$, and $\beta_0 = \arg \max_{\beta} \inf_y f(y)/f_0(y|\beta)$. We may interpret p_0 either as an upperbound of the probability $P(W = 0)$ that the ceteris paribus condition on the "state of the world" holds, or as the maximal Bayesian prior that the model $f_0(y|\beta)$ is correctly specified. In either case, we can estimate p_0 , and its estimate may serve as a measure of the information contents of a theoretical model $f_0(y|\beta)$ about the data-generating process $f(y)$, and hence as a measure of how realistic the theoretical model is. We call this probability, p_0 , the **reality bound** of the theoretical model² involved. In this paper we consider three versions of the reality bound: (1) the **marginal reality bound**, where $f(y)$ and $f_0(y|\beta)$ are the marginal densities of a single observation implied by the data-generating process and the theoretical model, respectively; (2) the **average conditional reality bound**, which is the average of the probabilities p_0 in the case that $f(y)$ and $f_0(y|\beta)$ are conditional densities; and (3) the **joint reality bound**, where $f(y)$ and $f_0(y|\beta)$ are the joint densities of the sample implied by the data-generating process and the theoretical model.

These reality bounds differ from other goodness of fit measures, for example Watson's (1993) measures of the fit of a real business cycle model, or the various approaches in Pagan (1994), in that they can be interpreted as probabilities that the theoretical model is correct, whereas other measures of fit are more akin to R^2 -type measures. Moreover, p_0 is an information measure³ rather than a goodness of fit measure: it measures how much information about the data-generating process is contained in the theoretical model.

The rest of the paper is organized as follows. Section 2 contains a brief historical perspective on the use of the ceteris paribus assumption in economics. In Section 3, the ceteris paribus assumption is described within the context of a fully stochastic theoretical economic model. As mentioned above, this is done by introducing a "state of the world" variable which may be interpreted either as an index of omitted variables, or as a random parameter. We show how the theoretical model is related to the data-generating process, by conditioning on the event that the "state of the world" variable is constant. In Section 3, we also outline the derivation of the various **reality bounds**. In Section 4, the theoretical tools introduced in the previous section are used to examine the consequences of the ceteris paribus assumption within the context of a linearized version of the real business cycle (RBC) model of King, Plosser, and Rebelo (1988b).

2 Some historical background on the ceteris paribus assumption

The specification, estimation, and testing of econometric models is closely linked to the construction of economic theories. While this statement holds for a number of rather obvious reasons, there is at least one link which can be interpreted very differently, depending upon whether one is a theorist or an empiricist. This link is the ceteris paribus⁴ assumption, which is common to both econometric (or statistical) and theoretical economic models, but which is in some respects much more crucial for econometric modelling than for the construction and interpretation of economic theories. In order to see the link between the use of the ceteris paribus assumption in economic theory and econometrics, it is useful to start by examining the approach taken by most theorists when constructing economic theories. Beliefs are first simplified into some parsimonious set of postulates and hypotheses. The simplest theories possible are then built up around these postulates and hypotheses, where by simplest theories possible, we mean that models are presented which are sufficiently simplistic enough to both convey and contain the essentials of the theory, as well as maintain enough generality as to be realistic. The latter of these two requirements (that of sufficient generality) is closely linked with the notion of ceteris paribus, the Latin expression for "other things being equal". One of the simplest theories which makes use of a ceteris paribus assumption is the Law of

Demand (i.e. quantity demanded depends negatively on price *ceteris paribus*, etc.). This is a very useful and convenient theory, as long as the *ceteris paribus* assumption is not ignored, and it is understood that complements as well as substitutes exist for most traded goods, for example. Indeed, the *ceteris paribus* assumption has been invoked throughout the history of economics, probably in large part because simpler theories more readily convey ideas, and also because without this assumption we would need a unified theory of economics, and just as a unified theory of physics has not been developed, and may not be developed for many generations, so a unified theory for economics also seems far away.

Examples of the use of the *ceteris paribus* assumption in economic theory date back more than one hundred years. For example, consider Beardsley (1895), where the effect of an eight hour workday on wages and unemployment is discussed. Beardsley notes that one popular theory of his day was that wages varied positively with the productiveness of an industry, and so shorter hours would reduce wages if they lessened production. Beardsley goes on to point out that such a statement is dependent critically on the *ceteris paribus* assumption, in particular:

Ceteris paribus, wages vary with the productiveness of industry, but only ceteris paribus. The theory that wages depend entirely on the efficiency of labor, or on the product of industry, is a new form of the old doctrine of the wages-fund. The characteristic feature of the classical doctrine was the assumption that the wages-fund was an inelastic quantum of the total circulating capital.

The notion of *ceteris paribus* has indeed been used in many historical papers to illustrate theoretical concepts (e.g. Edgeworth (1904), Murray et al. (1913), Pigou (1917)), and is still used in much current research (e.g. Lewis and Sappington (1992), Eisner (1992), and Gosh and Ostrey (1997)). In fact, a simple search of *The American Economic Review*, *Econometrica*, *The Journal of Economic History*, *The Journal of Industrial Economics*, *The Journal of Political Economy*, *The Quarterly Journal of Economics*, *The Review of Economics and Statistics*, *Journal of Applied Econometrics*, and *The Journal of Economic Perspectives* resulted in 2098 papers being found over the last 100 years which included the phrase *ceteris paribus*. Assuming 100 articles per year for 100 years, for each of these

approximately 10 journals suggests that approximately 2% of economics papers contain mention of the ceteris paribus assumption. Although this number is a very crude estimate, it highlights the importance of the assumption, particularly when it is noted that most uses of versions of the assumption do not explicitly refer to it as the ceteris paribus assumption.

While the ceteris paribus assumption, and the related concept of parsimony, are used in theory to help convey ideas, and to retain simplicity given a largely endogenous economy, these concepts must be viewed in a very different light when empirical models are being specified, estimated, and tested. Put another way, the ceteris paribus assumption can be easily understood in theoretical economic models, and in most cases is rather innocuous, as long as its potential importance is not overlooked. However, when the same ceteris paribus assumption is carried over into the fields of empirical model specification, estimation, and inference, careful note of its potential impact must be made. Indeed, the notion of ceteris paribus is so important in empirical econometrics that at least one very broad area in econometrics attempts in many ways to deal with its potential ramifications. This area involves the study of exogeneity in econometric models.

According to Koopmans (1950): '*A variable is exogenous if it is determined outside the system under analysis*'. Engle, Hendry, and Richard (1983) make this concept operational within the context of econometrics by formulating concepts of weak exogeneity, strong exogeneity, and super exogeneity, in which relationships between contemporaneous variables and parameters of interest are examined. Essentially, Engle et al. (1983) note that the ceteris paribus assumption (i.e., ignoring variables which may be "important") may have severe consequences on estimation and inference. For example, if the true economy is a complex endogenous system, then estimating a subsystem of the bigger endogenous system may result in an estimated model in which the estimated parameters are not consistent estimates of the true parameters of interest from the original theoretical model. These types of problems, which do not plague theoretical economics, have been well known for many years. For example, Keuzenkamp (1995) discusses the controversy between Keynes and Tinbergen on econometric testing of business cycle theories, and notes that multiple regression, which was sometimes thought to take care of required ceteris paribus conditions, does not actually help to counter Keynes view on Tinbergen's methods. Put in modern terminology, Keynes might have argued that Tinbergen's models suffered from a lack of weak exogeneity with

respect to the parameters of interest. Obviously, this was a difficult criticism for Tinbergen to counter, as the need for parsimony and tractability leads econometricians to use simple stochastic specifications, at least to some degree. However, Tinbergen was certainly aware of the issue. For example, Tinbergen (1935) notes that:

The aim of business cycle theory is to explain certain movements of economic variables. Therefore, the basic question to be answered is in what ways movements of variables may be generated. In answering this question it is useful to distinguish between exogen[ous] and endogen[ous] movements, the former being movements during which certain data vary, while, in the latter, the data are supposed to be constant. . . . We have now to sum up which groups of subjects contribute to the supply of the demand in each of the groups of markets and also how these supply and demand contributions behave in dependence on the variables adopted. Within certain limits, this choice of variables is free. We can state a priori, however, that categories remaining constant, or nearly constant, throughout a business cycle should not be taken as variables.

Thus, econometricians have grappled with the assumption of exogeneity, and the related assumption of ceteris paribus for many decades, and the paper by Engle et al. (1983) can be interpreted as a modern answer to the old problem. As another example of exogeneity and the importance of the ceteris paribus assumption, note that Gini (1937) states:

It is clear that, in order to measure the contribution of the different circumstances in the determination of a certain phenomenon, it is necessary not only to know their effects when each of them acts separately, but also to know the manner in which their effects combine when they act simultaneously.

This statement by Gini, put within the econometric context of Engle et al., says that we are generally interested in the joint density of a group of economic variables, as well as in the individual marginal densities. However, in practice, estimation is often carried out by factoring a joint density into a conditional and a marginal density, and weak exogeneity states that the parameters of interest can be consistently estimated from the conditional

density. Thus, for the purpose of estimation of the parameters of interest, and given an appropriately defined conditional density, the marginal density can essentially be ignored. This is clearly a form of the ceteris paribus assumption. In order to further clarify the link between ceteris paribus and exogeneity, note that Hendry, Pagan, and Sargan (1984) state that:

As a slight caricature, economic-theory based models require strong ceteris paribus assumptions (which need not be applicable to the relevant data generation process) and take the form of inclusion information such as $y = f(z)$, where z is a vector on which y is claimed to depend. While knowledge that z may be relevant is obviously valuable, it is usually unclear whether z may in practice be treated as "exogenous" and whether other variables are irrelevant or are simply assumed constant for analytical convenience (yet these distinctions are important for empirical modelling).

The main difference between the ceteris paribus assumption in economic theory and econometrics is that in economic theory the ceteris paribus condition is actually imposed, even if it is clear that the "state-of-the-world" is not fixed, whereas in econometrics it is often only used as a thought experiment which facilitates the interpretation of estimation results. The exogeneity issue is one example of the latter, where one of the questions which is addressed is: Under what conditions is the ceteris paribus assumption harmless for the estimation of the parameters of interest. Another example of a ceteris paribus thought experiment in econometrics is the interpretation of the coefficients of a linear regression model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$, say, as marginal effects: β_1 is the effect on the conditional expectation $E[y|x_1, x_2]$ of a unit increase in x_1 , given the ceteris paribus condition that x_2 is (held) constant. If x_1 and x_2 are independent, we could even impose this condition without affecting the consistency of the OLS estimate of β_1 . Of course, this is not the only possible thought experiment in this case. One can impose the ceteris paribus assumption in different ways and at different conditioning stages. For example, the linear regression model involved can be rewritten as $y = \beta_0 + \beta_1 x_1 + \beta_2 E[x_2|x_1] + \beta_2 (x_2 - E[x_2|x_1]) + u$, and now the effect on the conditional expectation $E[y|x_1, x_2]$ of a unit increase in x_1 , given the ceteris paribus condition that only the "innovation" $x_2 - E[x_2|x_1]$ is (held) constant, is the same as the effect on $E[y|x_1]$ of a unit increase in x_1 . See also Manski (1997, Section. 2.4) for similar ceteris

paribus thought experiments in the case of econometric models of response to treatment and covariates. In particular, Manski (1997) considers two types of ceteris paribus assumptions on the covariates, namely before and after the covariates themselves have responded to the treatment.

Finally, consider a quotation (taken from Loui (1989)) which is from the seminal book on modern decision theory by Savage (1950):

In application of the theory, the question will arise as to which [description of the] world to use. . . . If the person is interested in the only brown egg in a dozen, should that egg or the whole dozen be taken as the world? It will be seen . . . that in principle no harm is done by taking the larger of the two worlds as a model of the situation.

This statement summarizes succinctly our dilemma. We would like to examine a large portion of the world, and given the correct model specification, we should learn more by examining this large portion of the world rather than a smaller one. However, our models are always approximations, hence the more complex the model, the larger the likelihood of model misspecification. Moreover, we are left with the problem of determining whether our "portion" of the world is general enough to adequately mimic the characteristics of the economy in which we are interested. In this paper we address these problems by providing a measure of the extent of misspecification of theoretical economic models.

3 Relating theoretical economic models to the real world

3.1 The ceteris paribus condition on the "state of the world"

Let Y be a vector of dependent variables, and let X be a possibly infinite dimensional vector of exogenous and predetermined variables. Thus, X may include lagged values of Y in the case of time series models. For convenience, we assume in the discussion below that X is finite-dimensional so that we can condition on events of the form $X = x$, where x is a finite-dimensional non-random vector. Our arguments, however, extend to the case of an infinite dimensional X .

Assume that a fully stochastic theoretical economic model is the conditional density⁵ of

the "model version" of Y , given $X = x$,

$$f_{\text{mod}}(y|x, \beta), \tag{1}$$

where β is a parameter vector. Of course, this conditional density may not be equal to the actual conditional density, $f(y|x)$, of Y given $X = x$.

In order to show how model (1) is related to the actual conditional density $f(y|x)$, let us introduce a random variable or vector, W , which represents the "state of the world". There are two convenient interpretations of W which we will discuss. First, let W be a vector of omitted variables. If the omitted variables involved are countable-valued (which is always the case in practice), we can map these variables one-to-one onto the natural numbers [see Bierens (1988) and Bierens and Hartog (1988)], and hence we may (and will) assume that W is a discrete nonnegative random variable.

Alternatively, we may assume that W represents a discrete random parameter. This second interpretation leads to a Bayesian explanation of W , which we discuss in a later section.

Turning again to our discussion of how one can compare theoretical models with actual data-generating processes, let the true conditional density of Y given $X = x$ and $W = w$ be

$$f(y|x, w). \tag{2}$$

Note also that the *ceteris paribus* assumption can be imposed by assuming that W is a nonrandom scalar. Thus, and without loss of generality, we may assume that the economic theorist conditions on the event that $W = 0$. The theoretical model, (1), is therefore correctly specified, given the *ceteris paribus* condition involved, if:

$$f_{\text{mod}}(y|x, \beta) = f(y|X = x, W = 0), \tag{3}$$

for some parameter vector β . On the other hand, the conditional density of Y given only that $X = x$ is:

$$f(y|x) = \sum_{w=0}^{\infty} f(y|x, w)p(w|x), \tag{4}$$

where $p(w|x)$ is the conditional probability function of W , given $X = x$. Since we are only interested in the comparison of (3) with (4), we may assume that the "state of the world" variable W is a binary variable, because we can always write (4) as

$$f(y|x) = p(0|x)f_0(y|x) + (1 - p(0|x)) f_1(y|x), \quad (5)$$

where $f_0(y|x) = f(y|x, 0)$ and $f_1(y|x) = (1 - p(0|x))^{-1} \sum_{w=1}^{\infty} f(y|x, w)p(w|x)$.

3.2 The "state of the world" as an index of omitted variables

As an example of the interpretation of W as representing omitted variables, suppose that the data set is a random sample of workers, and that the theoretical model is a standard Mincer-type wage equation. Let Y be the log of wages, and X be a vector of characteristics of the worker involved, such as years of schooling, years of experience, years of experience squared, and a race dummy. Let W be a vector of omitted variables. For example, let W include higher powers and/or products of years of experience, regional dummy variables, dummy variables for various types of schooling, regional economic indicators, indicators of types of industries, and a dummy variable for marital status. For the sake of the argument, let us assume that the error term of the model including the omitted variables is distributed $N(0, \sigma_*^2)$, conditionally on X and W , and that the conditional expectation of Y given $X = x$ and $W = w$ is linear. Then, the true conditional density of Y given $X = x$ and $W = w$ takes the form:

$$f(y|x, w) = \frac{\exp\left(-\frac{1}{2}(y - \alpha_0 - \alpha'_1 x - \gamma' w)^2 / \sigma_*^2\right)}{\sigma_* \sqrt{2\pi}},$$

and the actual conditional density of Y given $X = x$ alone is:

$$f(y|x) = \int f(y|x, w) dP(w|x), \quad (6)$$

where $P(w|x)$ is the conditional distribution function of W given $X = x$. If we assume conditional normality of the error term, u , in the original Mincer-type wage equation as well, i.e. $u|X \sim N(0, \sigma^2)$, the conditional density implied by this model becomes:

$$f_{\text{mod}}(y|x, \beta) = \frac{\exp\left(-\frac{1}{2}(y - \beta_0 - \beta'_1 x)^2 / \sigma^2\right)}{\sigma \sqrt{2\pi}}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \sigma \end{pmatrix},$$

which in general differs from the actual conditional density $f(y|x)$, but is related to the conditional density $f(y|x, w)$. This can be seen by setting w equal to a constant vector, say $w = 0$, in which case:

$$f_{\text{mod}}(y|x, \beta) = f(y|X = x, W = 0),$$

for some value of β . Thus, the *ceteris paribus* condition involved is that $W = 0$, or equivalently, that W is constant.

In this example the vector W contains omitted variables that are discrete. Bierens and Hartog (1988) show that for such countable-valued random vectors almost any linear transformation $\theta'W$ is a one-to-one mapping (i.e. $W = \Phi_\theta(\theta'W)$ for some Borel measurable mapping Φ_θ), with the exception of θ 's in a set with zero Lebesgue measure. Thus, we may pick a suitable θ , and replace W by $\Phi_\theta(W_*)$, where $W_* = \theta'W$. Therefore, without loss of generality we may assume in this case that the "state of the world" variable W is a scalar random variable rather than a random vector. Moreover, given (5), we may further assume that W is a dummy variable.

3.3 The "state of the world" as a binary random parameter

We now show that under mild conditions the true conditional density, $f(y|x)$, can be written as (5). Thus, we may interpret W as a binary random parameter with prior (conditional) probability function $p(w|x)$, $w = 0, 1$. The main difference between this set-up and the usual Bayesian set-up⁶ is that the prior density involved does not represent the prior belief⁷ of the theorist in his model, but is constructed from the theoretical density, (3), and the true density, $f(y|x)$, as follows.

Consider two continuous distribution functions on \mathfrak{R} , say $F(y)$ and $F_0(y)$, with corresponding densities $f(y)$ and $f_0(y)$, respectively, where $F(y)$ is the true distribution function of some random variable Y , and $F_0(y)$ is the "model version" of $F(y)$. For convenience, we examine univariate unconditional distributions. However, our arguments also apply to multivariate (conditional) distributions. Our approach is to squeeze the distribution function $F_0(y)$ under the distribution function $F(y)$ such that for some number, $p_0 \in (0, 1)$, we can write $F(y) = p_0F_0(y) + (1 - p_0)F_1(y)$, where F_1 is a distribution function. This is possible if

we can find a positive p_0 such that $f(y) - p_0 f_0(y) \geq 0$ on the support of $f_0(y)$. The maximal p_0 for which this is possible is:

$$p_0 = \inf_{f_0(y) > 0} \frac{f(y)}{f_0(y)}.$$

Note that p_0 is nonnegative, and cannot exceed 1 if the supports of f_0 and f are the same, because in this special case $F(y) - p_0 F_0(y)$ is a nonnegative monotonic non-decreasing function with limit $1 - p_0 \geq 0$ for $y \rightarrow \infty$. If the support of $f(y)$ is contained in the support of $f_0(y)$, and if $f(y) = 0$ in an area where $f_0(y) > 0$, then $p_0 = 0$. In this case the theoretical model is able to predict impossible values for Y , and such models are logically inconsistent with reality. Thus, the result that $p_0 = 0$ is appropriate in this case. However, if the support of $f_0(y)$ is contained in the support of $f(y)$, and if $\int_{f_0(y) > 0} f(y) dy > 0$, then there is no guarantee that $p_0 \leq 1$. This is the case for real business cycle models, where the support of $f_0(y)$ is a lower-dimensional subspace of the support of $f(y)$. We shall deal with this case in the next subsection.

Given the above considerations, we assume for the remainder of this section that the supports of $f_0(y)$ and $f(y)$ are equal. However, even in this case, it is possible that $p_0 = 0$. For example, assume that $f(y)$ is the density of the standard normal distribution, and $f_0(y)$ is the density of the $N(0, \sigma^2)$ distribution with $\sigma^2 > 1$. Then

$$\inf_y f(y)/f_0(y) = \sigma \inf_y \exp \left[-(1/2)y^2 (1 - \sigma^{-2}) \right] = 0.$$

In practice, though, $f_0(y) = f_0(y|\beta)$ depends on parameters, and so does

$$p_0(\beta) = \inf_{f_0(y|\beta) > 0} f(y)/f_0(y|\beta).$$

Letting

$$p_0 = \sup_{\beta} p_0(\beta) \tag{7}$$

there will be a better chance that $p_0 > 0$. For example, if $f_0(y|\beta)$ is the density of the $N(0, \beta^2)$ distribution, and $f(y)$ is the density of the standard normal distribution, then

$$\begin{aligned} p_0(\beta) &= \inf_y f(y)/f_0(y|\beta) = |\beta| \inf_y \exp \left[-(1/2)y^2 (1 - \beta^{-2}) \right] \\ &= \begin{cases} 0 & \text{if } |\beta| > 1, \\ |\beta| & \text{if } |\beta| \leq 1, \end{cases} \end{aligned}$$

hence $p_0 = 1$. In any case we can write $F_1(y) = (F(y) - p_0 F_0(y)) / (1 - p_0)$, where $F_1(y) = F(y)$ if $p_0 = 0$. This distribution function is continuous itself, with density $f_1(y)$, hence:

$$f(y) = p_0 f_0(y) + (1 - p_0) f_1(y).$$

In the case that $f(y) = f(y|x)$, and $f_0(y) = f(y|x, W = 0)$ is parametrized as $f_{\text{mod}}(y|x, \beta)$, p_0 depends on x and β :

$$p_0(x|\beta) = \inf_{f_{\text{mod}}(y|x, \beta) > 0} \frac{f(y|x)}{f_{\text{mod}}(y|x, \beta)}, \quad (8)$$

so that (5) becomes

$$f(y|x) = p_0(x|\beta) f_{\text{mod}}(y|x, \beta) + (1 - p_0(x|\beta)) f_1(y|x, \beta).$$

Similarly to (7) we could take $p_0(x) = \sup_{\beta} p_0(x|\beta)$ as an upperbound of the conditional probability that $f_{\text{mod}}(y|x, \beta)$ is correct, but then $\beta_0(x) = \arg \max_{\beta} p_0(x|\beta)$ will depend on x . Therefore we propose the following "average" conditional reality bound:

$$p_0 = \sup_{\beta} E[p_0(X|\beta)].$$

Since both interpretations of W essentially yield the same result for the actual conditional density, namely equation (5), we shall also call W the "the state of the world" variable in the case where W is a random parameter.

3.4 The case of nested supports

The above Bayesian interpretation of W as a random parameter is particularly convenient in the case of stochastic general equilibrium models such as the real business cycle models advocated by Kydland and Prescott (1982) and their followers, because due to the highly stylized nature of these models and the single representative agent assumption it is not realistic to attribute their lack of fit entirely to the ceteris paribus assumption when it is equated with the presence of omitted variables in the theoretical model.

However, in the case of real business cycle models the support of the theoretical density is a lower-dimensional subspace of the support of the data-generating process. Thus, the approach in the previous section is not directly applicable. In this subsection we shall show

why this approach is not applicable in the case of standard real business cycle models, and in the next subsection we shall discuss an alternative and related approach.

To begin with, consider a simple example. Let the true density of the random vector $Y = (Y_1, Y_2)'$ be:

$$f(y) = f(y_1, y_2) = \frac{\exp\left[-\frac{1}{2}(y_1^2/\sigma_1^2 + y_2^2/\sigma_2^2)\right]}{\sigma_1\sigma_2 2\pi},$$

which is modelled as:

$$f_0(y|\beta) = f_0(y_1, y_2|\beta) = \frac{\exp\left(-\frac{1}{2}y_2^2/\beta^2\right)}{\beta\sqrt{2\pi}}I(y_1 = 0),$$

where $I(\cdot)$ is the indicator function. Thus, $f_0(y|\beta)$ is the density of the singular bivariate normal distribution:

$$N_2\left(0, \sigma^2 \begin{pmatrix} 0 & 0 \\ 0 & \beta^2 \end{pmatrix}\right).$$

Then, the support of $f_0(y|\beta)$ is the subspace spanned by the vector $(0, 1)'$. Therefore, we have that:

$$\begin{aligned} \inf_{f_0(y_1, y_2|\beta) > 0} \frac{f(y_1, y_2)}{f_0(y_1, y_2|\beta)} &= \inf_{y_2 \in \mathfrak{R}} \frac{f(0, y_2)}{f_0(0, y_2|\beta)} \\ &= \inf_{y_2 \in \mathfrak{R}} \frac{\beta}{\sigma_1\sigma_2\sqrt{2\pi}} \exp\left[\frac{1}{2}(\beta^{-2} - \sigma_2^{-2})y_2^2\right] \\ &= \begin{cases} 0 & \text{if } \beta^2 > \sigma_2^2 \\ \frac{\beta}{\sigma_1\sigma_2\sqrt{2\pi}} & \text{if } \beta^2 \leq \sigma_2^2. \end{cases} \end{aligned}$$

Hence,

$$\sup_{\beta} \inf_{f_0(y_1, y_2|\beta) > 0} \frac{f(y_1, y_2)}{f_0(y_1, y_2|\beta)} = \frac{1}{\sigma_1\sqrt{2\pi}},$$

which is larger than 1 if $\sigma_1 < \sqrt{2\pi}$.

The problem with the approach of the previous section arises because the theoretical model, $f_0(y_1, y_2|\beta)$, imposes the ceteris paribus condition ($Y_1 = 0$), which is not integrated out from $f(y_1, y_2)$. In other words, $f_0(y_1, y_2|\beta)$ is compared with the wrong data-generating process. The model density $f_0(y_1, y_2|\beta)$ in this example is actually the conditional density of Y_2 given the ceteris paribus condition $W = Y_1 = 0$. Hence, we should compare it with the marginal density:

$$f(y_2) = \int f(w, y_2)dw = \frac{\exp\left(-\frac{1}{2}y_2^2/\sigma_2^2\right)}{\sigma_2\sqrt{2\pi}},$$

rather than with $f(y_1, y_2)$ itself.

3.5 Reality bounds for singular normal models.

Now, consider the more general case where $f(y)$ is the density of a k -variate normal distribution, $N_k(\omega, -)$, where $-$ is nonsingular, and let the theoretical model $f_0(y|\beta)$ be the density of the k -variate singular normal distribution, $N_k(\mu(\beta), \Sigma(\beta))$, where $\text{rank}(\Sigma(\beta)) = m < k$. Also, assume that ω and $-$ are given. This case applies in particular to linearized versions of real business cycle models, which is also our area of application (in Section 4). Therefore we pay here explicit attention to the singular normal case.

It is difficult to write down the closed form of $f_0(y|\beta)$ on its support. Fortunately, there is no need for this, as the shapes of $f(y)$ and $f_0(y|\beta)$ are invariant under rotation and location shifts. Therefore, instead of working with $f(y)$ and $f_0(y|\beta)$ we may without loss of generality work with the transformed densities:

$$f_z(z|\beta) = f(\Pi(\beta)z + \mu(\beta))$$

and

$$f_{z,0}(z|\beta) = f_0(\Pi(\beta)z + \mu(\beta)|\beta),$$

respectively, where $\Pi(\beta)$ is the orthogonal matrix of eigenvectors of $\Sigma(\beta)$. Partitioning $\Pi(\beta) = (\Pi_1(\beta), \Pi_2(\beta))$, where $\Pi_1(\beta)$ is the $k \times (k - m)$ matrix of eigenvectors corresponding to the zero eigenvalues of $\Sigma(\beta)$, and $\Pi_2(\beta)$ is the $k \times m$ matrix of eigenvectors corresponding to the positive eigenvalues $\lambda_1(\beta), \dots, \lambda_m(\beta)$ of $\Sigma(\beta)$, we have that:

$$f_z(z|\beta) = f_z(z_1, z_2|\beta) = \frac{\exp\left[-\frac{1}{2}(\Pi(\beta)z - \tau(\beta))' -^{-1}(\Pi(\beta)z - \tau(\beta))\right]}{(\sqrt{2\pi})^k \sqrt{\det -}},$$

where

$$\tau(\beta) = \omega - \mu(\beta), \quad z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \in \Re^{k-m} \times \Re^m.$$

Also,

$$f_{z,0}(z|\beta) = f_{z,0}(z_1, z_2|\beta) = \frac{\exp\left(-\frac{1}{2}z_2' \Lambda(\beta)^{-1} z_2\right)}{(\sqrt{2\pi})^m \sqrt{\det \Lambda(\beta)}} I(z_1 = 0),$$

where

$$\Lambda(\beta) = \text{diag}(\lambda_1(\beta), \dots, \lambda_m(\beta)).$$

Again, the latter density is actually the conditional density of Z_2 given the ceteris paribus condition (that $W = Z_1 = 0$), and should therefore be compared with the marginal density:

$$\begin{aligned} f_z(z_2|\beta) &= \int f_z(z_1, z_2|\beta) dz_1 \\ &= \frac{\exp\left[-\frac{1}{2}(z_2 - \Pi_2(\beta)'\tau(\beta))'(\Pi_2(\beta)' - \Pi_2(\beta))^{-1}(z_2 - \Pi_2(\beta)'\tau(\beta))\right]}{(\sqrt{2\pi})^m \sqrt{\det(\Pi_2(\beta)' - \Pi_2(\beta))}}. \end{aligned}$$

Denoting $\pi_2(\beta) = \Pi_2(\beta)'\tau(\beta) = \Pi_2(\beta)'(\omega - \mu(\beta))$, we now have that:

$$\begin{aligned} p_0 &= \sup_{\beta} \inf_{z_2} \frac{f_z(z_2|\beta)}{f_{z,0}(0, z_2|\beta)} \\ &= \sup_{\beta} \sqrt{\det\left[\Lambda(\beta)^{1/2}(\Pi_2(\beta)' - \Pi_2(\beta))^{-1}\Lambda(\beta)^{1/2}\right]} \\ &\quad \times \exp\left[-\frac{1}{2}\pi_2(\beta)'(\Pi_2(\beta)' - \Pi_2(\beta))^{-1}\pi_2(\beta)\right] \\ &\quad \times \inf_{z_2} \left\{ \exp\left[\frac{1}{2}z_2'(\Lambda(\beta)^{-1} - (\Pi_2(\beta)' - \Pi_2(\beta))^{-1})z_2\right] \right. \\ &\quad \left. \times \exp\left[z_2'(\Pi_2(\beta)' - \Pi_2(\beta))^{-1}\pi_2(\beta)\right] \right\}. \end{aligned}$$

If the matrix $\Lambda(\beta)^{-1} - (\Pi_2(\beta)' - \Pi_2(\beta))^{-1}$ is positive definite, which is the case if

$$\lambda_{\max}\left[\Lambda(\beta)^{1/2}(\Pi_2(\beta)' - \Pi_2(\beta))^{-1}\Lambda(\beta)^{1/2}\right] < 1,$$

where $\lambda_{\max}[A]$ is the maximum eigenvalue of A , then

$$\begin{aligned} &\inf_{z_2} \left\{ \exp\left[\frac{1}{2}z_2'(\Lambda(\beta)^{-1} - (\Pi_2(\beta)' - \Pi_2(\beta))^{-1})z_2\right] \right. \\ &\quad \left. \times \exp\left[z_2'(\Pi_2(\beta)' - \Pi_2(\beta))^{-1}\pi_2(\beta)\right] \right\} \\ &= \exp\left[-\frac{1}{2}\pi_2(\beta)'(\Pi_2(\beta)' - \Pi_2(\beta))^{-1}(\Lambda(\beta)^{-1} - (\Pi_2(\beta)' - \Pi_2(\beta))^{-1})^{-1} \right. \\ &\quad \left. \times (\Pi_2(\beta)' - \Pi_2(\beta))^{-1}\pi_2(\beta)\right], \end{aligned}$$

so that

$$\begin{aligned} p_0 &= \sup_{\beta, \lambda_{\max}[\Psi(\beta)] < 1} \left\{ \sqrt{\det \Psi(\beta)} \exp\left[-\frac{1}{2}\vartheta(\beta)'\Psi(\beta)\vartheta(\beta)\right] \right. \\ &\quad \left. \times \exp\left[-\frac{1}{2}\vartheta(\beta)'\Psi(\beta)(I - \Psi(\beta))^{-1}\Psi(\beta)\vartheta(\beta)\right] \right\}, \end{aligned} \tag{9}$$

where

$$\Psi(\beta) = \Lambda(\beta)^{1/2}(\Pi_2(\beta)' - \Pi_2(\beta))^{-1}\Lambda(\beta)^{1/2} \tag{10}$$

and

$$\vartheta(\beta) = \Lambda(\beta)^{-1/2} \Pi_2(\beta)' (\omega - \mu(\beta)). \quad (11)$$

Clearly, $p_0 \leq 1$, because $\det \Psi(\beta) \leq \lambda_{\max}[\Psi(\beta)]$ if $\lambda_{\max}[\Psi(\beta)] < 1$.

In view of this argument, we may assume without loss of generality that the supports of the densities $f(y)$ and $f_0(y|\beta)$ are (or have been made) the same. This allows us to compute various versions of (8) for a variety of theoretical models.

3.6 The prior probability as a measure of the reality content of a theoretical model

In order to compute (8), we have to estimate the density $f(y|x)$. One way of doing this is by nonparametric estimation. However, nonparametric conditional density estimation requires large samples, which we usually do not have, particularly when we are examining macroeconomic data. Moreover, nonparametric density estimators suffer from the curse of dimensionality. Therefore, the only practical way to proceed is to specify a parametric functional form for $f(y|x)$, say $f(y|x, \theta)$, in general.

If we adopt our Bayesian interpretation of W , then in principle we can estimate the prior conditional probability, (8), with $f(y|x)$ replaced by $f(y|x, \hat{\theta})$, where $\hat{\theta}$ is the maximum likelihood estimator of θ based on the optimization of (8) with respect to β , so that:

$$\hat{p}_0(x) = \sup_{\beta \in B} \inf_{f_{\text{mod}}(y|x, \beta) > 0} \frac{f(y|x, \hat{\theta})}{f_{\text{mod}}(y|x, \beta)},$$

where B is the parameter space of the theoretical model. However, in this case:

$$\hat{\beta}_0(x) = \arg \max_{\beta \in B} \left(\inf_{f_{\text{mod}}(y|x, \beta) > 0} \frac{f(y|x, \hat{\theta})}{f_{\text{mod}}(y|x, \beta)} \right)$$

depends on x , which means that we estimate the parameter vector β for each $x = X_t$ separately! Clearly, this would paint too rosy a picture of the theoretical model. Therefore, in the conditional case we propose the following statistic:

$$\hat{p}_C = \sup_{\beta} \frac{1}{n} \sum_{t=1}^n p_0(X_t | \beta, \hat{\theta}), \quad (12)$$

where

$$p_0(x | \beta, \theta) = \inf_{f_{\text{mod}}(y|x, \beta) > 0} \frac{f(y|x, \theta)}{f_{\text{mod}}(y|x, \beta)}. \quad (13)$$

The rationale behind this proposal is that $p_0(x|\beta, \theta)$ is a conditional probability. Therefore, $p_0(\beta, \theta) = E [p_0(X_t|\beta, \theta)]$ is the corresponding unconditional probability, so that \hat{p}_C is an estimate of the maximal unconditional probability, $p_0(\theta) = \sup_{\beta} p_0(\beta, \theta)$. In other words, we may interpret \hat{p}_C as an estimate of the maximum average probability that the conditional model $f_{\text{mod}}(y|x, \beta)$ is correctly specified, and we therefore call \hat{p}_C the estimated *average conditional reality bound*.

We may also replace the conditional densities in (8) with either the joint densities of the data or the marginal densities of a single observation, Y_t , in order to get rid of x . In particular, let $Y = (Y'_1, \dots, Y'_n)'$ be a vector of stacked variables, and let the joint density of Y implied by the theoretical model be:

$$f_{\text{mod}}(y|\beta), \quad \beta \in B,$$

Also, let the functional specification of the true joint density of Y be:

$$f(y|\theta), \quad \theta \in \Theta,$$

so that $L_{\text{mod}}(\beta) = f_{\text{mod}}(Y|\beta)$ is the likelihood implied by the theoretical model, and $L(\theta) = f(Y|\theta)$ is the likelihood function of the data-generating process. Moreover, let $\hat{\theta}$ be the maximum likelihood estimator of θ . Then:

$$\hat{p}_J = \sup_{\beta \in B} \inf_{f_{\text{mod}}(y|\beta) > 0} \frac{f(y|\hat{\theta})}{f_{\text{mod}}(y|\beta)}, \quad (14)$$

is an estimate of the probability that $W = 0$, which may be interpreted as the maximum probability that the theoretical joint density, $f_{\text{mod}}(y|\beta)$, is correctly specified. Therefore, \hat{p}_J also serves as a measure of the reality content of the theoretical model, and the larger is \hat{p}_J , the more realistic is the theoretical economic model. We call \hat{p}_J the estimated *joint reality bound*.

The computation of (14) in the case where f and f_{mod} represent the joint densities of the sample turns out to be a formidable numerical problem. However, if f and f_{mod} represent only the marginal densities of a single Y_t , then the computation of (14) is quite feasible. In this case we will denote the estimated probability, (14), as:

$$\hat{p}_M = \sup_{\beta \in B} \inf_{f_{\text{mod}}(y|\beta) > 0} \frac{f(y|\hat{\theta})}{f_{\text{mod}}(y|\beta)}, \quad (15)$$

where f and f_{mod} are now marginal densities. We call \hat{p}_M the estimated *marginal reality bound*. In the empirical application below, we construct only \hat{p}_C and \hat{p}_M , as the preferred reality bound, \hat{p}_J , is too difficult to compute, due to the singularity of the real business cycle model involved.

4 Measuring the marginal and average conditional reality bounds of a real business cycle model

4.1 The model

As an illustration of our approach, consider the baseline real business cycle (RBC) model of King, Plosser and Rebelo (KPR: 1988b), which is derived from Kydland and Prescott (1982), and which is used in Watson (1993)⁸:

$$\max E_0 \sum_{t=0}^{\infty} \lambda^t (\ln(C_t) + \theta \ln(1 - N_t)), \quad \lambda < 1,$$

subject to

$$Q_t = C_t + I_t = K_t^{1-\alpha} (A_t N_t)^\alpha,$$

$$\ln(A_t) = \ln \gamma_x + \ln(A_{t-1}) + \varepsilon_t, \quad \varepsilon_t \text{ i.i.d. } (0, \sigma^2), \quad (16)$$

$$K_t = (1 - \delta)K_{t-1} + I_t,$$

where C_t denotes consumption, Q_t is output, K_t is capital, N_t is employment, I_t is investment, and A_t is a technology index. Note that technology follows a random walk with drift equal to $\ln \gamma_x$. Except for A_t , the variables in the above theoretical model may all be interpreted as per capita aggregates. In order to examine the empirical predictions of this model, KPR (1988a,b) and Stock (1993) calculate the equilibrium of the model using the log linearization of the Euler equations proposed by KPR (1987). Using this approximate solution, it turns out that the theoretical model can be written as a restricted VAR(1)⁹,

$$Y_t = A(\beta)Y_{t-1} + \mu(\beta) + u_t, \quad u_t \text{ i.i.d. } (0, \Sigma(\beta)), \quad (17)$$

where

$$Y_t = \begin{pmatrix} \Delta \ln(Q_t) \\ \Delta \ln(C_t) \\ \Delta \ln(I_t) \\ \ln(N_t) \end{pmatrix}, \quad (18)$$

with $\beta = (\lambda, \theta, \alpha, \delta, \gamma_x, \sigma^2)'$, and $A(\beta)$ a matrix with elements depending on β . Augmenting model (17) with the assumption that u_t is normally distributed,

$$u_t \sim N(0, \Sigma(\beta)),$$

the joint distribution of the data, and the marginal distribution of Y_t implied by the theoretical model, are completely determined. The functional form of $\mu(\beta)$, $A(\beta)$, and $\Sigma(\beta)$ will be discussed below.

Following KPR (1988a,b) and Watson (1993), we parameterize the true data-generating process as a stationary Gaussian VAR(p):

$$Y_t = \xi + \sum_{i=1}^p \Xi_i Y_{t-i} + v_t, \quad v_t \sim i.i.d. N_4(0, \Gamma). \quad (19)$$

Stacking all of the parameters involved in a vector, $\theta \in \Theta$, the marginal density $f(y | \theta)$ is

$$f(y | \theta) = \frac{\exp \left[-\frac{1}{2} (y - \omega(\theta))' - (\theta)^{-1} (y - \omega(\theta)) \right]}{(2\pi)^2 \sqrt{\det - (\theta)}}, \quad (20)$$

where $\omega(\theta) = \xi + \sum_{i=1}^{\infty} \Psi_i \xi$ and $- (\theta) = \Gamma + \sum_{i=1}^{\infty} \Psi_i \Gamma \Psi_i'$, with

$$\left(I - \sum_{i=1}^p \Xi_i L^i \right)^{-1} = I + \sum_{i=1}^{\infty} \Psi_i L^i.$$

Note that the randomness of the theoretical model version of Y_t is solely due to the random variable, ε_t , in (16). Thus, the variance matrix of the theoretical marginal model ($\Sigma(\beta)$) is singular.

4.2 VAR representation

It follows directly from equations (2.3)-(2.8) in KPR (1988b) that the theoretical VAR model (17), augmented with the normality assumption takes the form:

$$\Delta \ln(Q_t) = \ln(\gamma_x) - \frac{\pi_{QK}}{\pi_{NK}} (\phi - 1) \ln(n) + \frac{\pi_{QK}}{\pi_{NK}} (\phi - 1) \ln(N_{t-1}) \quad (21)$$

$$\begin{aligned}
& + \left(\frac{1}{\pi_{NK}} - \frac{\pi_{QK}}{\pi_{NK}} \right) \varepsilon_t, \\
\Delta \ln(C_t) &= \ln(\gamma_x) - \frac{\pi_{CK}}{\pi_{NK}}(\phi - 1) \ln(n) + \frac{\pi_{CK}}{\pi_{NK}}(\phi - 1) \ln(N_{t-1}) \\
& + \left(\frac{1}{\pi_{NK}} - \frac{\pi_{CK}}{\pi_{NK}} \right) \varepsilon_t, \\
\Delta \ln(I_t) &= \ln(\gamma_x) - \frac{\pi_{IK}}{\pi_{NK}}(\phi - 1) \ln(n) + \frac{\pi_{IK}}{\pi_{NK}}(\phi - 1) \ln(N_{t-1}) \\
& + \left(\frac{1}{\pi_{NK}} - \frac{\pi_{IK}}{\pi_{NK}} \right) \varepsilon_t \\
\ln(N_t) &= (1 - \phi) \ln(n) + \phi \ln(N_{t-1}) + \varepsilon_t, \quad \varepsilon_t \sim i.i.d. N(0, \sigma^2),
\end{aligned}$$

where $\ln N_t - \ln n$ is the percent deviation of N_t from its steady state path, with $\ln n$ constant (recall that the solution method which KPR (1988b) use to solve for the equilibrium of the theoretical model amounts to linearizing the first-order conditions of the choice problem about the steady state, after transforming the economy to obtain a stationary system). The parameter ϕ above can be interpreted in a number of ways. For example, KPR (1988b) note that when ϕ is close to unity, hours will be characterized by substantial serial correlation, while the growth rates of Q_t , C_t , and I_t will exhibit little serial correlation. Thus, when ϕ is close to unity an investigator might be tempted to difference the log of hours in addition to the logs of Q_t , C_t , and I_t , which would lead to misspecification. Finally, note that π_{QK} , π_{CK} , π_{IK} , and π_{NK} should be interpreted as the elasticities of the flow variables with respect to deviations of the capital stock from its stationary value¹⁰.

4.3 The marginal reality bound

We can write this theoretical model as:

$$Y_t = \mu + AY_{t-1} + \kappa\varepsilon_t, \quad \varepsilon_t \sim i.i.d. N(0, \sigma^2), \quad (22)$$

where $A = (O, \alpha)$ is a 4×4 matrix with α a 4×1 vector, O is a 4×3 zero matrix, and μ and κ 4×1 vectors. The relationship between the vectors α , μ , and κ is:

$$\begin{aligned}
\mu &= \ln(\gamma_x) \iota_{13} + \ln(n) (\iota_4 - \alpha), \\
\kappa &= \frac{1}{\pi_{NK}} \iota_{13} + \frac{1 - 2\phi}{1 - \phi} \iota_4 + \frac{1}{1 - \phi} \alpha
\end{aligned} \quad (23)$$

$$\begin{aligned}
&= \frac{1}{\pi_{NK}} \iota_{13} + \frac{1 - 2\alpha_4}{1 - \alpha_4} \iota_4 + \frac{1}{1 - \alpha_4} \alpha \\
&= \frac{1}{\pi_{NK}} \iota_{13} + \frac{1 - 2\alpha_4}{1 - \alpha_4} \iota_4 + \frac{\alpha_4}{1 - \alpha_4} \beta_{13} \\
&= \beta_4 \beta_{13} + (1 - \beta_4) \iota_4 + \beta_5 \iota_{13}, \text{ and}
\end{aligned}$$

$$\alpha = \alpha_4 \beta_{13} = \frac{\beta_4}{1 + \beta_4} \beta_{13},$$

where

$$\iota_{13} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \quad \iota_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad \beta_{13} = \begin{pmatrix} \alpha_1/\alpha_4 \\ \alpha_2/\alpha_4 \\ \alpha_3/\alpha_4 \\ 1 \end{pmatrix}, \quad \beta_4 = \frac{\alpha_4}{1 - \alpha_4} > 0, \quad \beta_5 = \frac{1}{\pi_{NK}},$$

and $\alpha_4 = \phi$ is the last component of α . The corresponding MA(∞) representation is:

$$Y_t = \sum_{j=0}^{\infty} A^j \mu + \sum_{j=0}^{\infty} A^j \kappa \varepsilon_{t-j},$$

where $A^0 = I$. Using the facts that $A^2 = \alpha_4 A$, with α_4 the last component of α , and $\mu_4 = (1 - \alpha_4) \ln(n)$ is the last component of μ , it follows that

$$\begin{aligned}
E_{\text{mod}}(Y_t) &= \mu + A\mu \left(\sum_{j=1}^{\infty} \alpha_4^{j-1} \right) = \mu + \frac{\mu_4}{1 - \alpha_4} \alpha = \mu + \ln(n) \alpha \\
&= \ln(\gamma_x) \iota_{13} + \ln(n) \iota_4 = \beta_6 \iota_{13} + \beta_7 \iota_4,
\end{aligned}$$

where

$$\beta_6 = \ln(\gamma_x), \quad \beta_7 = \ln(n).$$

Moreover, since the last component of κ is 1, so that $A\kappa = \alpha$, it follows that:

$$\begin{aligned}
\frac{\text{Var}_{\text{mod}}(Y_t)}{\sigma^2} &= \kappa \kappa' + \frac{1}{1 - \alpha_4^2} A \kappa \kappa' A' = \kappa \kappa' + \frac{1}{1 - \alpha_4^2} \alpha \alpha' \\
&= \kappa \kappa' + \frac{\alpha_4^2}{1 - \alpha_4^2} \beta_{13} \beta_{13}' = \kappa \kappa' + \frac{\beta_4^2}{1 + 2\beta_4} \beta_{13} \beta_{13}' = C(\beta),
\end{aligned} \tag{24}$$

say. Clearly, the rank of this matrix cannot exceed 2, and it has rank 1 only if κ is proportional to β_{13} , which is not the case. Therefore, the rank of $C(\beta)$ is 2.

It is not hard to verify that the two positive eigenvalues of the matrix $C(\beta)$ are:

$$\begin{aligned}\lambda_1(\beta) &= \frac{1}{2} \left(\kappa' \kappa + \frac{\beta'_{13} \beta_{13} \beta_4^2}{1 + 2\beta_4} \right) \\ &\quad - \frac{1}{2} \sqrt{\left(\kappa' \kappa - \frac{\beta'_{13} \beta_{13} \beta_4^2}{1 + 2\beta_4} \right)^2 + \frac{4(\kappa' \beta_{13})^2 \beta_4^2}{1 + 2\beta_4}}, \\ \lambda_2(\beta) &= \frac{1}{2} \left(\kappa' \kappa + \frac{\beta'_{13} \beta_{13} \beta_4^2}{1 + 2\beta_4} \right) \\ &\quad + \frac{1}{2} \sqrt{\left(\kappa' \kappa - \frac{\beta'_{13} \beta_{13} \beta_4^2}{1 + 2\beta_4} \right)^2 + \frac{4(\kappa' \beta_{13})^2 \beta_4^2}{1 + 2\beta_4}},\end{aligned}\tag{25}$$

where

$$\kappa' \kappa = \beta_4^2 \beta'_{13} \beta_{13} + 2\beta_4 \beta_5 \beta'_{13} \beta_{13} - \beta_4^2 + 3\beta_5^2 + 1,$$

and

$$\kappa' \beta_{13} = \beta_4 \beta'_{13} \beta_{13} + \beta_5 \beta'_{13} \beta_{13} - \beta_4 + 1.$$

Moreover, the corresponding eigenvectors, $q_j(\beta)$, are:

$$q_j(\beta) = \begin{cases} \frac{(\kappa' \beta_{13}) \kappa - (\kappa' \kappa - \lambda_j(\beta)) \beta_{13}}{\|(\kappa' \beta_{13}) \kappa - (\kappa' \kappa - \lambda_j(\beta)) \beta_{13}\|} & \text{if } \kappa' \beta_{13} \neq 0, \\ \frac{\kappa}{\|\kappa\|} & \text{if } \kappa' \beta_{13} = 0 \text{ and } \lambda_j(\beta) = \kappa' \kappa, \\ \frac{\beta_{13}}{\|\beta_{13}\|} & \text{if } \kappa' \beta_{13} = 0 \text{ and } \lambda_j(\beta) = \frac{\beta_4^2}{1+2\beta_4} \beta'_{13} \beta_{13}. \end{cases}$$

We now have all the ingredients needed in order to compute \hat{p}_M . In (9), let:

$$\Lambda(\beta) = \sigma^2 \begin{pmatrix} \lambda_1(\beta) & 0 \\ 0 & \lambda_2(\beta) \end{pmatrix} = \sigma^2 \Lambda_*(\beta),$$

$$\Pi_2(\beta) = (q_1(\beta), q_2(\beta)),$$

$$\Psi(\beta) = \sigma^2 \Lambda_*(\beta)^{1/2} \left(\Pi_2(\beta)' - (\hat{\theta}) \Pi_2(\beta) \right)^{-1} \Lambda_*(\beta)^{1/2} = \sigma^2 \Psi_*(\beta),$$

and

$$\vartheta(\beta) = \sigma^{-1} \Lambda_*(\beta)^{-1/2} \Pi_2(\beta)' \left(\omega(\hat{\theta}) - \beta_6 \beta_{13} - \beta_7 \beta_4 \right) = \sigma^{-1} \vartheta_*(\beta).$$

Note that $\sqrt{\det \Psi(\beta)} = \sigma^2 \sqrt{\det \Psi_*(\beta)}$, and that the condition $\lambda_{\max}[\Psi(\beta)] < 1$ now becomes:

$$\sigma^2 < \frac{1}{\lambda_{\max}[\Psi_*(\beta)]}.$$

It follows from (9) that:

$$\begin{aligned} \hat{p}_M &= \sup_{\beta|\sigma^2 < 1/\lambda_{\max}[\Psi_*(\beta)]} \left\{ \sigma^2 \sqrt{\det \Psi_*(\beta)} \exp \left[-\frac{1}{2} \vartheta_*(\beta)' \Psi_*(\beta) \vartheta_*(\beta) \right] \right. \\ &\quad \left. \times \exp \left[-\frac{1}{2} \sigma^2 \vartheta_*(\beta)' \Psi_*(\beta) \left(I - \sigma^2 \Psi_*(\beta) \right)^{-1} \Psi_*(\beta) \vartheta_*(\beta) \right], \right. \end{aligned} \quad (26)$$

where

$$\Psi_*(\beta) = \Lambda_*(\beta)^{1/2} \left(\Pi_2(\beta)' - (\hat{\theta}) \Pi_2(\beta) \right)^{-1} \Lambda_*(\beta)^{1/2}$$

and

$$\vartheta_*(\beta) = \Lambda_*(\beta)^{-1/2} \Pi_2(\beta)' \left(\omega(\hat{\theta}) - \beta_6 \iota_{13} - \beta_7 \iota_4 \right). \quad (27)$$

Given β , we can concentrate out σ^2 by maximizing the right-hand side of (26) (over the interval $(0, 1/\lambda_{\max}[\Psi_*(\beta)])$) with respect to σ^2 by simple line search, and then optimize with respect to β . However, in the current context, it turns out that:

$$\hat{p}_M = 1. \quad (28)$$

The reason for this is twofold. First, observe that (27) can be made equal to zero by choosing β_6 and β_7 , given β_1, \dots, β_5 , such that:

$$\Pi_2(\beta)' \omega(\hat{\theta}) = \beta_6 \Pi_2(\beta)' \iota_{13} + \beta_7 \Pi_2(\beta)' \iota_4 = \Pi_2(\beta)' (\iota_{13}, \iota_4) \begin{pmatrix} \beta_6 \\ \beta_7 \end{pmatrix},$$

which is a system of two equations in two unknowns, β_6 and β_7 . Therefore, we may set $\vartheta_*(\beta) = 0$ and $\sigma^2 = \lambda_{\max}[\Psi_*(\beta)]$. Then, it follows that:

$$\hat{p}_M = \sup_{\beta} \frac{\sqrt{\det \Psi_*(\beta)}}{\lambda_{\max}[\Psi_*(\beta)]} = \sup_{\beta} \frac{\sqrt{\lambda_{\min}[\Psi_*(\beta)] \lambda_{\max}[\Psi_*(\beta)]}}{\lambda_{\max}[\Psi_*(\beta)]}, \quad (29)$$

where $\lambda_{\min}[\Psi_*(\beta)]$ is the minimum eigenvalue of $\Psi_*(\beta)$. Clearly, \hat{p}_M is equal to unity if $\lambda_{\min}[\Psi_*(\beta)] = \lambda_{\max}[\Psi_*(\beta)]$, which is a single nonlinear equation in 5 unknowns, β_1, \dots, β_5 , and can be solved. (We verified that $\hat{p}_M = 1$, using numerical methods.)

However, rather than concluding that the baseline model is perfect, it appears that the comparison of marginal distributions in this case is not very informative. Nevertheless, the inclusion of the marginal reality bound in our discussion is useful as an illustration of the general method, and because the conclusion (28) is not obvious.

4.4 The average conditional reality bound

The theoretical conditional distribution of Y_t is:

$$Y_t|X_t \sim N_4 \left(\mu_t(\beta), \sigma^2 \kappa \kappa' \right),$$

where

$$X_t = (Y'_{t-1}, \dots, Y'_{t-p})'$$

and

$$\begin{aligned} \mu_t(\beta) &= \mu + AY_{t-1} = \mu + \alpha \iota'_4 Y_{t-1} = \ln(\gamma_x) \iota_{13} + \ln(n)(\iota_4 - \alpha) + \alpha \iota'_4 Y_{t-1} \\ &= \ln(\gamma_x) \iota_{13} + \ln(n) \iota_4 + (\iota'_4 Y_{t-1} - \ln(n)) \alpha \\ &= \beta_6 \iota_{13} + \beta_7 \iota_4 + \beta_4 (\iota'_4 Y_{t-1} - \beta_7) \beta_{13}, \end{aligned}$$

using the same reparametrization and notation as above. The DGP is:

$$Y_t|X_t \sim N_4 (\omega_t(\theta), -(\theta)),$$

where $\omega_t(\theta) = \xi + \sum_{i=1}^p \Xi_i Y_{t-i}$ and $-(\theta) = \Gamma$.

The matrix $\kappa \kappa'$ has one positive eigenvalue ($\lambda(\beta) = \kappa' \kappa$), and corresponding normalized eigenvector, $q(\beta) = \kappa / \|\kappa\|$. Note, however, that we can write $\mu_t(\beta)$ in terms of κ and β_j , $j = 4, \dots, 7$ as $\mu_t(\beta) = \beta_6 \iota_{13} + \beta_7 \iota_4 + (\iota'_4 Y_{t-1} - \beta_7) (\kappa - \iota_4 + \beta_4 \iota_4 - \beta_5 \iota_{13})$. Therefore, without loss of generality, we may set $\beta_{13} = \kappa$, so that:

$$\mu_t(\beta) = \beta_6 \iota_{13} + \beta_7 \iota_4 + (\iota'_4 Y_{t-1} - \beta_7) (\beta_{13} - \iota_4 + \beta_4 \iota_4 - \beta_5 \iota_{13}),$$

$$\lambda(\beta) = \beta'_{13} \beta_{13} \quad \text{and} \quad q(\beta) = \beta_{13} / \sqrt{\lambda(\beta)}.$$

Now, $\Pi_2(\beta)$ and $\Lambda(\beta)$ in (9) become $q(\beta)$ and $\sigma^2 \lambda(\beta)$, respectively, and $\vartheta(\beta)$ becomes

$$\vartheta_t(\beta) = \frac{q(\beta)' (\omega_t(\theta) - \mu_t(\beta))}{\sigma \sqrt{\lambda(\beta)}} = \frac{\beta'_{13} (\omega_t(\theta) - \mu_t(\beta))}{\sigma \beta'_{13} \beta_{13}}.$$

Also, $\Psi(\beta)$, together with the condition $\lambda_{\max}[\Psi(\beta)] < 1$ becomes:

$$\Psi(\beta) = \frac{\sigma^2 (\beta'_{13} \beta_{13})^2}{\beta'_{13} (\widehat{\theta}) \beta_{13}} < 1.$$

Moreover, $\beta'_{13}\mu_t(\beta)$ can be reparametrized as:

$$\beta'_{13}\mu_t(\beta) = \beta_4 + \beta_5 l'_4 Y_{t-1},$$

and $\Psi(\beta)$ can be reparametrized as $\beta_6^2 = \Psi(\beta) < 1$. Thus, we can now write (13) according to (9) as:

$$p_0(X_t|\beta, \hat{\theta}) = \beta_6 \exp \left[-\frac{1}{2} \left(\frac{1}{1 - \beta_6^2} \right) \frac{(\beta'_{13}\omega_t(\hat{\theta}) - \beta_4 - \beta_5 l'_4 Y_{t-1})^2}{\beta'_{13} (\hat{\theta}) \beta_{13}} \right]. \quad (30)$$

Finally,

$$\hat{p}_C = \sup_{\beta|0 < \beta_6 < 1} \frac{1}{n} \sum_{t=1}^n p_0(X_t|\beta, \hat{\theta}),$$

which can be computed by numerical methods.

Note how crucial the restrictions on the conditional mean of the theoretical model are. If it were possible to choose β such that $\beta'_{13}\omega_t(\hat{\theta}) = \beta'_{13}\mu_t(\beta)$ for all t , then we could set $\beta_6 = 1$, so that $\hat{p}_C = 1$.

Finally, note from Jensen's inequality that:

$$\hat{p}_C = \exp \left[\sup_{\beta|0 < \beta_6 < 1} \frac{1}{n} \sum_{t=1}^n \ln(p_0(X_t|\beta, \hat{\theta})) \right] \leq \hat{p}_C.$$

This lowerbound can be written as:

$$\begin{aligned} \hat{p}_C &= \frac{1}{2} \sup_{\beta|0 < \beta_6 < 1} \left\{ \ln(\beta_6^2) - \frac{\eta(\beta)}{1 - \beta_6^2} \right\} \\ &= \frac{1}{2} \left\{ \sup_{\beta} \ln \left(1 + \frac{1}{2}\eta(\beta) - \sqrt{\left(1 + \frac{1}{2}\eta(\beta)\right)^2 - 1} \right) \right. \\ &\quad \left. - \frac{\eta(\beta)}{1 - \frac{1}{2}\eta(\beta) + \sqrt{\left(1 + \frac{1}{2}\eta(\beta)\right)^2 - 1}} \right\}, \end{aligned} \quad (31)$$

where

$$\eta(\beta) = \frac{\frac{1}{n} \sum_{t=1}^n (\beta'_{13}\omega_t(\hat{\theta}) - \beta_4 - \beta_5 l'_4 Y_{t-1})^2}{\beta'_{13} (\hat{\theta}) \beta_{13}},$$

and is much easier to calculate than \hat{p}_C . Thus, the "optimal" β 's obtained from the construction of \hat{p}_C may be used as starting values in the numerical computation of \hat{p}_C .

4.5 Empirical result

In the previous subsection we have outlined the tools with which to examine the conditional baseline RBC model of KPR (1988b). We now use these tools to construct an estimate of the average conditional reality bound (ACRB). We recall that our measure is an information measure: it measures the extent of information about the data-generating process that is contained in the theoretical model. It can be viewed as an alternative to the use of other methods, such as Euler equation restriction tests, moment matching, impulse response analysis, duration analysis, and spectral analysis (see e.g. all of the articles in the special issue of the *Journal of Applied Econometrics* on calibration edited by Pagan (1994)). The advantage of our approach is that we obtain a single probability value which summarizes the ability of the RBC model to match the data, relative to some more general model where the *ceteris paribus* assumption has not been imposed.

Our approach is perhaps closest to the approach taken by Watson (1993), where the variables in the theoretical model are augmented with just enough stochastic error so that the model can exactly match the second moments of the actual data. Measures of fit for the model, called relative mean square approximation errors (RMSAEs), are then constructed on the basis of the variance of this stochastic error relative to the variance of the actual series. In at least two respects, our approach is similar to Watson's. First, both approaches attempt to obtain "overall" measures of closeness of model distribution and data-generating process. Second, and more importantly, both approaches assume that the theoretical model is only an approximation of the true underlying model, and as such do not rely on a null hypothesis that the theoretical model is equal to the true model (see e.g. Watson (1993), p. 1015), as do standard goodness of fit measures. The importance of this point is perhaps best illustrated by Sargent (1998), where it is noted that viewing econometric models as approximations of the truth rather than assuming that the models are correctly specified (under the null) can lead to interesting results. For example, when all models are viewed as approximations, Sargent shows that simple "adaptive" forecasting techniques based on rolling regressions, where parameters are updated at each point in time, actually yield forecasts which are comparable to those based on "optimal" rational expectations theories. This result is perhaps surprising, given that the rolling regression approach is certainly not optimal in a standard

utility optimizing representative agent framework.

The data which we use correspond to those used by Watson (1993), although our sample period is updated through 1994:4, so that we have 188 quarterly U.S. observations from 1948:1 to 1994:4. As mentioned above, the variables which we model are the natural logarithm of hours worked ($\ln(N_t)$) and the growth rates of output ($\Delta \ln(Q_t)$), consumption ($\Delta \ln(C_t)$), and investment ($\Delta \ln(I_t)$). In order to make operational our measure of the ACRB and the LACRB, we need to construct $\omega_t(\hat{\theta})$ and ${}_t\hat{\theta}$, the conditional mean and variance of $Y_t|X_t$, where $Y_t = (\Delta \ln(Q_t), \Delta \ln(C_t), \Delta \ln(I_t), \ln(N_t))'$ and X_t is as defined above. (It should perhaps be noted that we do not at any stage use calibration techniques in our calculation of \hat{p}_C .) Before forming $\omega_t(\hat{\theta})$ and ${}_t\hat{\theta}$, however, we first undertook a specification search for our *true* VAR model. To begin with, we conducted unit root tests, and found that $\ln(Q_t)$, $\ln(C_t)$, and $\ln(I_t)$ are all I(1), so that the differences of these variables are all I(0). Also, $\ln(N_t)$ is I(0), at least at a 10% level. (In this analysis, we used augmented Dickey-Fuller and Phillips-Perron unit root tests). We then chose the lag order for our VAR model of Y_t by examining values of the Akaike and Schwarz information criteria for lag orders from 1 to 12. Based on the Schwarz criterion, one lag was preferred, while the Akaike criterion picked five lags. We also conducted a series of Johansen trace tests to determine the rank of the cointegrating space (for the sub-system consisting of our three I(1) variables, say $Z_t = (\ln(Q_t), \ln(C_t), \ln(I_t))'$). In particular, five different cointegration tests, corresponding to the five different deterministic trending assumptions discussed in Johansen (1995), were conducted. The results based on these tests were rather mixed. As an example of our findings, consider the two cointegration tests which assume a linear deterministic trend in the data (the first includes an intercept in the cointegrating relation, while the second includes an intercept and a trend in the cointegrating relation). For a VAR(1) in Z_t , the cointegrating ranks found based on these two tests were 3 and 1, respectively, while for a VAR(5) in Z_t , the ranks were 0 and 0. (Complete details of all of our empirical findings are available upon request from the authors.) Given our rather mixed empirical findings, we chose to specify the *true* model for Y_t as a VAR(5) with no cointegrating restrictions¹¹.

Our approach was to first construct the lower average conditional reality bound (LACRB). The associated parameter estimates from this exercise were then used to construct the ACRB. The reason why we adopted this approach is that there are six different parameters, $\beta_{13} =$

$(\beta_1, \beta_2, \beta_3, 1)'$, β_4 , β_5 , and β_6 , to jointly solve for in the optimization of $\frac{1}{n} \sum_{t=1}^n p_0(X_t|\beta, \hat{\theta})$.¹² In our empirical illustration, joint optimization with respect to so many parameters results in a probability surface with a number of "flats", so that when "reasonable" starting values are not used, standard numerical optimization procedures (such as the Broyden, Fletcher, Goldfarb, and Shanno quasi-Newton method which we use) result in parameter estimates which do not "move" from their starting values. Indeed, this was what occurred when we chose starting values which were not equal to those associated with the LACRB. As noted above, construction of starting values based on the LACRB is quite simple¹³. Given these starting values associated with the LACRB, we obtained convergence to the ACRB after less than 50 iterations, and the bound¹⁴ was:

$$\hat{p}_C = 0.77.$$

The ACRB is rather high, and suggests that the reality content of the baseline RBC model is perhaps higher than one might imagine, given that the RBC version of the *true* VAR model is so severely restricted. For example, it is surprising that the use of only one source of stochastic fluctuation in the RBC model yields such a high probability. Indeed, it appears that the very simple RBC model which we examine provides a rather effective approximation of our *true* model. Watson (1993, Table 1) finds RMSAEs (which can be interpreted like 1 minus the R^2 statistic from a regression based on the theoretical model) which range from as low as 0.21 to as high as 1.00, depending upon which variable is examined. Interestingly, our finding corresponds closely with the values (from approximately 0.20 to 0.30) found by Watson when the variable for which he minimizes the stochastic error (see above) is the same as the variable for which he constructs the RMSAE.

References

- Beardsley, C., Jr., 1895, The effect of an eight hours' day on wages and the unemployed, *Quarterly Journal of Economics*, 9, 450-459.
- Bickel, P.J. and K.A. Doksum, 1977, *Mathematical Statistics* (Prentice Hall, Englewood Cliffs, NJ)
- Bierens, H.J., 1988, ARMA memory index modeling of economic time series (with discussion), *Econometric Theory* 4, 35-59.
- Bierens, H.J., and J. Hartog, 1988, Non-linear regression with discrete explanatory variables, with an application to the earnings function, *Journal of Econometrics* 38, 269-299.
- DeJong, D.N., B.F. Ingram and C.H. Whiteman, 1996, A Bayesian approach to calibration, *Journal of Business & Economic Statistics* 14, 1-9.
- Edgeworth, F.Y., 1904, The theory of distribution, *The Quarterly Journal of Economics*, 18, 159-219.
- Eisner, R., 1992, Deficits: which, how much, and so what?, *The American Economic Review* 82, 295-298.
- Engle, R.F., D.F. Hendry, and J.-F. Richard, 1983, Exogeneity, *Econometrica* 51, 277-304.
- Ghosh, A.R. and J.D. Ostry, 1997, Macroeconomic uncertainty, precautionary saving, and the current account, *Journal of Monetary Economics* 40, 121-139.
- Gini, C., 1937, Methods of eliminating the influence of several groups of factors, *Econometrica* 5, 57-73.
- Hansen, L.P., and J.J. Heckman, 1996, The empirical foundations of calibration, *Journal of Economic Perspectives* 10, 87-104.
- Harvey, A., 1990, *The Econometric Analysis of Time Series*, 2nd Edition (Philip Allan, Hemel Hempstead).
- Hendry, D.F., A.R. Pagan, and J.D. Sargan, 1984, Dynamic specification, in Z. Grilliches and M. Intriligator, eds., *Handbook of Econometrics*, Volume 2 (North Holland, Amsterdam) Chapter 18.
- Horowitz, J.L., and C.F. Manski, 1995, Identification and robustness with contaminated and corrupted data, *Econometrica* 63, 281-302.
- Johansen, S., 1995, *Likelihood Based Inference in Cointegrated Vector Autoregressive*

Models (Oxford University Press, Oxford).

Keuzenkamp H.A., 1995, Keynes and the logic of econometric method, discussion paper 113 (CentER, Tilburg University, the Netherlands).

King, R.G., C.I. Plosser, and S.T. Rebelo, 1987, Production, growth and business cycles: Technical appendix, manuscript (University of Rochester).

King, R.G., C.I. Plosser, and S.T. Rebelo, 1988a, Production, growth and business cycles: I. The basic neoclassical model, *Journal of Monetary Economics* 21, 195-232.

King, R.G., C.I. Plosser, and S.T. Rebelo, 1988b, Production, growth and business cycles: II. New directions, *Journal of Monetary Economics* 21, 309-341.

King, R.G., C.I. Plosser, J.H. Stock, and M.W. Watson, 1991, Stochastic trends and economic fluctuations, *American Economic Review*, 81, 819-840.

Koopmans, T.C., 1950, When is an equation system complete for statistical purposes?, *Oxford Bulletin of Economics and Statistics* 54, 325-348.

Kullback, L., and R.A. Leibler, 1951, On information and sufficiency, *Annals of Mathematical Statistics* 22, 79-86.

Kydland, F.E, and E.C. Prescott, 1982, Time to build and aggregate fluctuations, *Econometrica* 50, 1345-1370

Kydland, F.E, and E.C. Prescott, 1996, The computational experiment: An econometric tool, *Journal of Economic Perspectives* 10, 69-85.

Lewis, T. and D.E.M. Sappington, 1992, Incentives for conservation and quality-improvement by public utilities, *American Economic Review* 82, 1321-1340.

Loui, R.P., 1989, Two heuristic functions for utility, *Proceedings AAAI Spring Symposium on Limited Rationality*, 81-86.

Manski, C.F., 1997, Monotone treatment response, *Econometrica* 65, 1311-1344.

Murray, N.C., A.C. Whitaker, W.C. Fisher, O.M.W. Sprague, B.M. Anderson, Jr., R.R. Bowker, E.W. Kemmerer, and S. Fisher, 1913, Standardizing the dollar - Discussion, *American Economic Review*, 3, 29-51.

Pagan, A. (ed.), 1994, Calibration Techniques and Econometrics, Special Issue of the *Journal of Applied Econometrics*, 9, Supplement.

Pigou, A.C., 1917, The value of money, *Quarterly Journal of Economics*, 32, 38-65.

- Sargent, T.J., 1998, The conquest of American inflation, working paper (Hoover Institution and University of Chicago).
- Savage, L., 1950, Foundations of Statistics (Dover).
- Sims, C.A., 1980, Macroeconomics and reality, *Econometrica* 48, 1-48.
- Sims, C.A., 1996, Macroeconomics and methodology, *Journal of Economic Perspectives* 10, 105-120.
- Swanson, N.R. and C.W.J. Granger, 1997, Impulse response functions based on a causal approach to residual orthogonalization in vector autoregressions, *Journal of the American Statistical Association* 92, 357-367.
- Tinbergen, J., 1935, Annual survey: Suggestions on quantitative business cycle theory, *Econometrica* 3, 241-308.
- Watson, M.W., 1993, Measures of fit for calibrated models, *Journal of Political Economy* 101, 1011-1041.
- White, H., 1994, Estimation, Inference, and Specification Analysis (Cambridge University Press, Cambridge).
- Working, E.J., 1927, What do statistical demand curves show?, *Quarterly Journal of Economics* 41, 212-235.

FOOTNOTES:

* *Corresponding author*: Herman J. Bierens, Department of Economics, Pennsylvania State University, University Park, PA, 16802 (email: hbierens@psu.edu). This paper was presented by the first author at the conference on *Principles of Econometrics* in Madison, Wisconsin, on May 1-2, 1998. The useful comments of Arthur Goldberger, Charles Manski, and Neil Wallace, on earlier versions of this paper are gratefully acknowledged. Swanson thanks the National Science Foundation (grant SBR-9730102) for research support.

¹ This approach is (somewhat) related to the statistical literature on mixtures, and contaminated sampling. See Horowitz and Manski (1995) and the references therein for the latter. As is well known, under fairly general conditions there exist many sequences of densities $f_j(y)$ such that for a wide class of densities $f(y)$ we can write $f(y) = \lim_{m \rightarrow \infty} \sum_{j=0}^m p_{m,j} f_j(y)$, where $p_{m,j} > 0$, and $\sum_{j=0}^m p_{m,j} = 1$. The problem addressed in this literature is then to determine classes of densities $f_j(y)$ for which $f(y)$ can be approximated well by the finite mixture $\sum_{j=0}^m p_{m,j} f_j(y)$. Kernel density estimates are examples of this kind of mixtures. Moreover, denoting $p_0 = \lim_{m \rightarrow \infty} p_{m,0}$, and $f_1^*(y) = \lim_{m \rightarrow \infty} \sum_{j=1}^m p_{m,j} f_j(y) / (1 - p_0)$, we have $f(y) = p_0 f_0(y) + (1 - p_0) f_1^*(y)$. In the contaminated sampling literature, $f_1^*(y)$ may be interpreted as the contamination of the unknown density $f_0(y)$, occurring with probability $1 - p_0$. The main difference with the statistical literature involved is that in our case both $f(y)$ and $f_0(y)$ are given, possibly up to an unknown parameter vector, and p_0 is constructed out of $f(y)$ and $f_0(y)$.

² A more appropriate name for p_0 would be the "maximum likelihood" of the theoretical model, but that will likely cause confusion.

³ Note that p_0 is related to the Kullback-Leibler (1951) Information Criterion $KLIC(\beta) = \int_{f_0(y|\beta) > 0} f_0(y|\beta) \ln[f_0(y|\beta)/f(y)] dy$ by the inequality $\inf_{\beta} KLIC(\beta) \leq \ln[1/p_0]$. Moreover, $\ln[1/p_0]$ as a measure of the discrepancy of $f_0(y|\beta)$ from $f(y)$ satisfies the same information inequality as the $KLIC$: $\ln[1/p_0] \geq 0$, and $\ln[1/p_0] = 0$ if and only if for some β_0 the set $\{y: f_0(y|\beta_0) \neq f(y)\}$ has Lebesgue measure zero. Cf. White (1994), p. 9.

⁴ According to Webster's dictionary, the correct pronunciation of the "c" in "ceteris paribus" is as the "c" in "classic". However, in European academic circles this "c" is often pronounced as either the "ch" in "church", or the first "c" in "cycle". It appears that the Webster pronunciation is in accordance with classical Latin spoken by the Romans, and the

latter two with dialects of "church Latin", which was the academic lingua franca for more than 1000 years in Europe.

⁵ Another class of theoretical models consists of conditional expectation models of the form $E_{\text{mod}}[Y|X = x, \beta] = \int y f_{\text{mod}}(y|x, \beta) dy$, where only the left-hand side is specified by the theoretical model. The approach in this paper, however, does not apply to this class of partially specified stochastic models.

⁶ See for example DeJong, Ingram and Whiteman (1996) for a Bayesian analysis of calibrated models.

⁷ We realize that in the eyes of a true Bayesian our Bayesian interpretation may be considered blasphemy.

⁸ Although our example assumes that momentary utility is additively separable, it should be noted that KPR (1988a,b) also consider the case where momentary utility is multiplicatively separable. However, the restricted vector autoregression which we use to characterize the RBC model is the same, regardless of which utility specification we use. For further details, including restrictions on momentary utility, the reader is referred to KPR (1988a,b).

⁹ See Watson (1993) for further discussion of this model, and Sims (1980), Swanson and Granger (1997) and the citations therein, for a discussion of VAR models.

¹⁰ In this version of the theoretical model, there are 8 parameters, including σ^2 . Therefore, there are 2 implicit restrictions among these parameters. However, we shall ignore these two restrictions in order to keep our analysis tractable.

¹¹ It should be noted that when we examined individual pairs of our three I(1) variables using unit cointegrating vector restrictions, we found some evidence of cointegration, as one might expect, given the empirical findings of King, Plosser, Stock, and Watson (1991), for example. These results suggest that even our *true* model should perhaps only be viewed as an approximation to the truth, as the failure of simpler pairwise cointegration results to match up with our system cointegration results may be accounted for by model misspecification. However, for the purpose of constructing our *true* model, we found that little explanatory power was added by including cointegrating restrictions of any form, and thus we assume that our simpler specification is an adequate representation of the truth.

¹² As mentioned above, there is also one parameter restriction, namely that $0 < \beta_6 < 1$. Thus, optimization is carried out by optimizing $\frac{1}{n} \sum_{t=1}^n p_0(X_t|\beta, \hat{\theta})$ with respect to β_{13} , β_4 ,

and β_5 , given a grid of different β_6 values, where the grid is incremented by some small value.

¹³ First, one solves the generalized eigenvalue problem $|C_1 - C_2\lambda| = 0$, where $C_1 = W'MW$, $C_2 = -^{-1}(\hat{\theta})$, $W = (\omega'_2\beta_{13}, \dots, \omega'_n\beta_{13})'$, ω_i , $i=2,3,\dots,n$, are the elements of $\omega_t(\hat{\theta})$, where $\omega(\theta)$ is defined in Section 4.1, $\hat{\theta}$ is the maximum likelihood estimator of θ , and where $M = I_{n-1} - X(X'X)^{-1}X'$, and X is an $(n-1) \times 2$ matrix with first column equal to an $(n-1) \times 1$ vector of ones, and second column equal to $(i'_4Y_1, \dots, i'_4Y_{n-1})'$. In order to do this, set $C = L^{-1}AL'^{-1}$, where $C_2 = LL'$, and solve the standard eigenvalue problem, $Cx = \lambda x$. Then the eigenvector, say x_{min} , associated with the minimal eigenvalue, say λ_{min} , can be normalized to obtain $\hat{\beta}_{13}$, whose last element is unity. Once $\hat{\beta}_{13}$ is known, $\eta(\hat{\beta}) = \lambda_{min}$ can be plugged into equation (31).

¹⁴ This bound was found to be the same when the increment used in the grid search for β_6 was varied between 0.01 and 0.0001.