Testing Intertemporal Rational Expectations Model with State Uncertainty: An Application to the Permanent Income Hypothesis

Chao-Hsi Huang
Department of Economics
National Tsing Hua University

Yue-Lieh Huang
Department of Quantitative Finance
National Tsing Hua University

Chung-Ming Kuan
Institute of Economics
Academia Sinica

This version: March 20, 2004

Author for correspondence: Chao-Hsi Huang, Department of Economics, National Tsing Hua University, Taipei 115, Taiwan; ckuan@econ.sinica.edu.tw.

† We would like to thank C.-F. Chung and C.-S. Mao for useful comments on early version of this paper. C.-M. Kuan gratefully acknowledges the research support from the National Science Council of the Republic of China (NSC89-2415-H-001-071). All remaining errors are ours.

Abstract

In this paper we take a different modeling approach based on the component driven (CD) model developed in Kuan, Huang, and Tsay (2003) to test the permanent income hypothesis (PIH), an example of intertemporal choice models. A key feature of this approach is that it explicitly allows for state uncertainty. By assuming that the labor income follows a CD process, we show that the agent's perception on the likelihoods of income innovations being permanent and transitory plays a crucial role in determining the optimal forecasts on the change of consumption. In particular, the effect of a current innovation is a weighted average of two distinct effects (resulted from permanent and transitory innovations), with the weights being the perceived likelihoods of respective states. Also, past innovations may affect consumption when there is a revision on the perceived likelihoods of previous states. If there is no state uncertainty, our result reduces to that of an existing model. Our empirical study shows that, while the CD model can characterize the U.S. consumption data well, the estimation results do not agree with the predictions of the PIH.

JEL classification: C22, C51, D91, E21

Keywords: component driven model, intertemporal choice model, permanent income hypothesis, permanent innovation, state uncertainty, transitory innovation

1 Introduction

One important development in the macroeconomic and finance literature in the past fifty years has been the gaining popularity in the use of intertemporal choice as the basic framework for theoretical development. The models based on the intertemporal choice framework cover numerous macro-finance topics. The prominent examples include theories on consumption (e.g., Modigliani and Brumberg, 1954; Friedman, 1957), investment (e.g., Lucas, 1967; Abel and Blanchard, 1983), labor supply and demand (e.g., Lucas and Rapping, 1969; Heckman, 1974; Sargent, 1978), government fiscal policy (e.g., Barro, 1979; Frenkel and Razin, 1996), current account determination (e.g., Bruno, 1976; Sachs, 1981; Razin, 1995), business cycles (e.g., Barro, 1981; Lucas, 1981), asset pricing (e.g., Lucas, 1978; Breeden, 1979), and dividend policy (e.g., Lintner, 1956), to name a few. One common feature of these models is that the reaction of the choice (or "endogenous") variable to permanent shocks is rather different from that to transitory ones. For instance, according to the permanent income-consumption model of Friedman (1957), a permanent innovation in income would exert a much larger effect on consumption than does a transitory one. The equilibrium business cycles model of Barro (1981), on the other hand, suggests that a transitory innovation in government purchase has a much larger effect on the country's real interest rate.

Many researchers have empirically explored the above mentioned feature of the intertemporal choice models. Such studies include, among others, Hall and Mishkin (1982) and Falk and Lee (1998) on the permanent income consumption model, Sahasakul (1986) on the model of optimal tax policy, Glick and Rogoff (1992) on the intertemporal current account model, Barro (1981) and Ahmed (1986) on the equilibrium business cycles model, Cochrane (1994) and Lee (1995) on the asset pricing model, and Lee (1996) on the dividend smoothing model. In these studies, it is typical to assume that a rational agent is able to know with perfect certainty the state (that is, permanence vs. transitoriness) as well as the magnitude of the innovations. This assumption, though quite convenient for deriving the optimal forecast of the variable of interest, may not be so realistic as one would like.

In the real world, it seems more reasonable to postulate that an agent does not have full knowledge about the innovation state. When state uncertainty is present, the agent must draw inferences on the state based on available information and may subsequently revise these inferences when new information suggests so. For example, a rational agent may revise downward the perception on the permanence of a past pay raise once he/she encounters an unexpected pay cut. Yet this perception may be revised upward when the agent does not experience any pay cut for an extended period of time. The inferences on the state as well as the subsequent revisions ought to have profound impacts on the agent's expectations and resulting behaviors. Failing to account for such state uncertainty may result in misleading inference regarding the validity of the intertemporal model examined.

To address the problem of state uncertainty, we propose using the component driven model (the CD model), recently developed in Kuan, Huang, and Tsay (2003, henceforth KHT), and demonstrate how such a modeling strategy can be applied to testing intertemporal models. Specifically, the proposed CD model contains a unit-root component and a stationary component; whether a particular component will be activated depends on an unobservable state variable whose law of motion is governed by certain probability laws. Contrary to existing models, the effects of innovations in the CD model are not fixed at all times but may be permanent or transitory in different time periods. Since this modeling approach allows the innovation states to alternate from time to time, it is able to highlight the uncertainty the agent faces when judging the current and past states of the variable's innovations.

In this paper, we specialize on testing the permanent income hypothesis (PIH), a typical example of intertemporal models. By assuming that the labor income follows a CD process, we derive the optimal forecasts on the change of consumption. It is shown that the agent's perception on the likelihoods of income innovations being permanent and transitory plays a crucial role in determining these forecasts. On one hand, the effect of a current innovation is a weighted average of two distinct effects (one resulted from a permanent innovation and the other due to a transitory innovation), with the weights being the perceived likelihoods of respective states. On the other hand, past innovations may also affect the change of consumption when there is a revision on the perceived likelihoods of previous states. In particular, if there is no state uncertainty, there would be no need to draw inferences on states, so that the optimal forecast on the change of consumption reduces to that of an existing model, such as Deaton (1987) or Flavin (1981). Our result shows that, without the unrealistic assumption of known innovation state, the CD model provides a more general framework under which a rational agent can utilize available information optimally in his/her expectations formation. This approach is thus more in line with the spirit of the rational expectations hypothesis.

The rest of the paper is organized as the following. In section 2, we introduce the CD model under the permanent income framework. In section 3, we show how the expectations about future incomes are formed when the labor income follows a CD process. It can be seen that the uncertainty about the state of income innovation as well as the revision on the perception of the past innovation states may affect the agent's expectations about future incomes and hence consumption. In section 4, we discuss model estimation and hypothesis testing on the implications of the PIH. The empirical analysis of U.S. consumption is presented in section 5. It is found that, while the proposed model can characterize the U.S. consumption data well, the estimation results do not agree with the predictions of the PIH. Section 6 concludes the paper.

2 The Component-Driven Model of Income

There are numerous intertemporal choice models. To show how state uncertainty may affect the forecast of intertemporal models, we focus on the PIH in this paper. In the empirical literature of the PIH (e.g., Campbell, 1987), the representative agent's real consumption c_t is usually assumed to follow the permanent income y_t^p, which is defined as the annuity value of the sum of real net wealth w_t and the expected present value of current and future labor income y_{t+i} $(i = 0, 1, 2, \ldots)$:

$$c_t = y_t^p = r\left[w_t + \frac{r}{1+r}\sum_{i=0}^{\infty}(1+r)^{-i}\,\mathbb{E}_t y_{t+i}\right], \tag{1}$$

where r is the real interest rate and $\mathbb{E}_t = \mathbb{E}\big(\,\cdot \mid \boldsymbol{\Omega}^t\big)$ denotes the mathematical expectation conditional on $\boldsymbol{\Omega}^t$, the individual's information set available at time t. Given the permanent income-consumption relation stipulated in equation (1), it is straightforward to show that the change in consumption $(\Delta c_t = c_t - c_{t-1})$ is the annuity value of the revisions in the expected labor income:[1]

$$\Delta c_t = \frac{r}{1+r}\sum_{i=0}^{\infty}(1+r)^{-i}\big(\mathbb{E}_t - \mathbb{E}_{t-1}\big)y_{t+i}. \tag{2}$$

Note that equations (1) and (2) are both related to the expected future labor income.

[1]Here, the law of motion of the real net wealth is postulated as $w_{t+1} = (1+r)w_t + y_t - c_t$.

To test the PIH, it is typical to set up a model for labor income and derive its optimal forecast accordingly. The commonly used models for labor income include the trend-stationary model in which the innovations all have transitory effects (e.g., Flavin, 1981), the difference-stationary model in which the innovations all have permanent effects (e.g., West, 1988), and the model that admits both permanent and transitory innovations in each period (e.g., Hall and Mishkin, 1982; Quah, 1990; Falk and Lee, 1998; Elwood, 1998). In these models, the (transitory vis à vis permanent) state of income innovations is assumed to be known a priori. This assumption facilitates the derivation of the optimal forecast but may not be very realistic. For example, Beaudry and Koop (1993) and Bradley and Jansen (1997) find that the effects of innovations are likely to change from time to time. There is, however, no room for the agent to revise his/her forecast in existing models even when the perceived state differs from that originally postulated in the model. As such, the income forecasts generated from these models may contain systematic errors and need not be optimal. This, in turn, may lead to false inference on individual's consumption behavior.

Alternatively, we can employ a more flexible model in which the innovation state is unknown to the agent. To this end, we propose using a variant of KHT's CD model for the labor income process so that state uncertainty is explicitly allowed. Let s_t denote an unobservable, random state variable taking the value of one or zero and $\{\upsilon_t\}$ be a white noise with mean zero and variance σ_υ^2. Our CD model of the labor income y_t is the sum of two components: $y_t = y_{1,t}^* + y_{0,t}^*$, such that

$$\begin{aligned} \Gamma(B)\Delta y_{1,t}^* &= \alpha_0 + s_t\upsilon_t, \\ \Psi(B)y_{0,t}^* &= (1-s_t)\upsilon_t, \end{aligned} \tag{3}$$

where $\Gamma(B) = 1 - \gamma_1 B - \cdots - \gamma_n B^n$ and $\Psi(B) = 1 - \psi_1 B - \cdots - \psi_m B^m$ are polynomials of the back-shift operator B, both with all the roots outside the unit circle. It is readily seen that the first component $y_{1,t}^*$ essentially follows an ARIMA$(n, 1, 0)$ model, while the second compoent $y_{0,t}^*$ is a stationary AR(m) model. Comparing with the CD model originally considered by KHT, the model (3) admits more general short-run dynamics in both components. This model will be referred to as a CD$(n, 1; m)$ model, signifying that it is a mixture of an ARIMA$(n, 1, 0)$ model and an AR(m) model.

In model (3), only one component is activated at each time, depending on the realization of the state variable. When $s_t = 1$, the first component $y_{1,t}^*$ is excited by υ_t, while

$y^*_{0,t}$ evolves without this innovation. As long as $s_t = 1$, y_t would behave like a unit-root process with drift, and v_t has a permanent effect on future labor income. When $s_t = 0$, $y^*_{0,t}$ is excited by v_t, but $y^*_{1,t}$ grows along a linear trend without the new innovation. In this case, y_t would behave like a trend-stationary process when $s_t = 0$, and v_t exerts only a transitory effect on future income. This model thus permits both difference-stationary and trend-stationary dynamics in different periods, and its innovations may have permanent or transitory effects. In particular, when $s_t = 1$ ($s_t = 0$) with probability one for all t, (3) reduces to the conventional difference-stationary (trend-stationary) model.

It is straightforward to show that the CD$(n, 1; m)$ model has an ARIMA representation with random MA coefficients:

$$\Gamma(B)\Psi(B)\Delta y_t = \alpha_0 \Psi(1) + \sum_{i=1}^{\kappa+1} \xi_{i,s_{t-i}} v_{t-i} + v_t, \tag{4}$$

where $\kappa = \max\{m, n\}$,

$$\xi_{1,s_{t-1}} = \begin{cases} -\psi_1, & \text{if } s_{t-1} = 1, \\ -1 - \gamma_1, & \text{otherwise,} \end{cases} \qquad \xi_{i,s_{t-i}} = \begin{cases} -\psi_i, & \text{if } s_{t-i} = 1, \\ \gamma_{i-1} - \gamma_i, & \text{otherwise,} \end{cases}$$

for $i = 2, \ldots, \kappa$, and the last coefficient is

$$\xi_{\kappa+1,s_{t-\kappa-1}} = \begin{cases} 0, & \text{if } s_{t-\kappa-1} = 1, \\ \gamma_\kappa, & \text{otherwise;} \end{cases}$$

$\psi_i = 0$ for $i > m$ and $\varphi_i = 0$ for $i > n$. From (4) we can see that the realization of the past states s_{t-i} determines the effects of past income innovations v_{t-i} on the current value of Δy_t. Yet the effects of past innovations are uncertain to the agent because s_{t-i} are not observable. It can also be seen that the current state s_t plays no role in determining Δy_t because $s_t v_t$ and $(1 - s_t)v_t$ are both present in (3), so that their joint effect does not depend on s_t. Intuitively, whether an innovation has a permanent or transitory effect should not be known to the agent at the time it occurs; the effect of an innovation can only be revealed in subsequent periods. Indeed, (4) suggests that the information of the state of v_{t-i} is embedded in future labor incomes: $y_{t-i+1}, y_{t-i+2}, \ldots$.

To illustrate the proposed model, we consider a special case of (3) that consists of an ARIMA$(1, 1, 0)$ component with $\alpha_0 = 0$ and a stationary AR(1) component. Table 1 gives the moving-average representation of this CD(1,1;1) model for $t = 1, \ldots, 7$ and the realization $\{s_1, \ldots, s_6\} = \{0, 0, 0, 1, 1, 0\}$. We set $y_t = 0$ for $t = 0, -1, \ldots$. Here, y_t is a

Table 1: A moving-average representation of the CD(1,1;1) model.

$y_t =$	$y^*_{1,t}$	$+\ y^*_{0,t}$
$y_1 =$	0	$+\ \upsilon_1$
$y_2 =$	0	$+\ \psi_1\upsilon_1 + \upsilon_2$
$y_3 =$	0	$+\ \psi_1^2\upsilon_1 + \psi_1\upsilon_2 + \upsilon_3$
$y_4 =$	υ_4	$+\ \psi_1^3\upsilon_1 + \psi_1^2\upsilon_2 + \psi_1\upsilon_3$
$y_5 =$	$(1+\gamma_1)\upsilon_4 + \upsilon_5$	$+\ \psi_1^4\upsilon_1 + \psi_1^3\upsilon_2 + \psi_1^2\upsilon_3$
$y_6 =$	$(1+\sum_{i=1}^2\gamma_1^i)\upsilon_4 + (1+\gamma_1)\upsilon_5$	$+\ \psi_1^5\upsilon_1 + \psi_1^4\upsilon_2 + \psi_1^3\upsilon_3 + \upsilon_6$
$y_7 =$	$(1+\sum_{i=1}^3\gamma_1^i)\upsilon_4 + (1+\sum_{i=1}^2\gamma_1^i)\upsilon_5 + s_7\upsilon_7$	$+\ \psi_1^6\upsilon_1 + \psi_1^5\upsilon_2 + \psi_1^4\upsilon_3 + \psi_1\upsilon_6 + (1-s_7)\upsilon_7$
	$\vdots$	$\vdots$
$y_\infty =$	$\upsilon_4/(1-\gamma_1)\ +\ \upsilon_5/(1-\gamma_1)\ +\ \cdots$	$+\ 0$

stationary AR(1) process at the beginning and starts evolving like an ARIMA process when $s_t = 1$. Given $0 < \gamma_1 < 1$ and $0 < \psi_1 < 1$, the impacts of $\upsilon_1, \upsilon_2, \upsilon_3$ and υ_6 on future incomes decay exponentially over time, but the impacts of υ_4 and υ_5 accumulate and converge to $\upsilon_4/(1-\gamma_1)$ and $\upsilon_5/(1-\gamma_1)$, respectively. Note that the realization of s_7 does not affect y_7. When s_t are not observable, the representation in Table 1 is just one possible sample path, and the true effect of an innovation on future incomes is unknown. Nonetheless, the information available at time t $(y_t, y_{t-1}, \ldots, y_{t-i+1})$ would be helpful in identifying the state of the past innovation υ_{t-i}.

3 Permanent Income-Consumption Relation

Given the CD income process in (3) and the permanent income-consumption relation in (1), we are now ready to derive the resulting consumption formula. A key feature of the CD process is that its sample path depends on the innovation states, but these states are not observable. Given state uncertainty, a rational agent must draw inferences on the state of current innovation and may revise his/her previous inferences when new information becomes available. It is thus reasonable to expect that the inferences on the state and subsequent revisions should affect the permanent income-consumption relation.

It is shown in Appendix I that, using the Wiener-Kolmogorov prediction formula discussed in Hansen and Sargent (1980, 1981), the consumption change Δc_t in (2) is

$$\begin{aligned}\Delta c_t &= \Gamma^* \,\mathbb{P}\big(s_t = 1 \mid \boldsymbol{\Omega}^t\big) \upsilon_t + \frac{r}{1+r} \Psi^* \,\mathbb{P}\big(s_t = 0 \mid \boldsymbol{\Omega}^t\big) \upsilon_t \\ &\quad + \Gamma^* \sum_{i=1}^{\infty} \gamma_i^* \zeta_{1,t-i} \upsilon_{t-i} + \frac{r}{1+r} \Psi^* \sum_{i=1}^{\infty} \psi_i^* \zeta_{0,t-i} \upsilon_{t-i},\end{aligned} \tag{5}$$

where $\boldsymbol{\Omega}^t = \{\Delta y_t, \Delta y_{t-1}, \ldots, \upsilon_t, \upsilon_{t-1}, \ldots\}$ is the information available up to time t, Γ^* and Ψ^* are the polynomials $\Gamma^{-1}(B)$ and $\Psi^{-1}(B)$ evaluated at the discount factor $1/(1+r)$, γ_i^* and ψ_i^* are the polynomial coefficients described in Appendix I. In this expression, $\mathbb{P}\big(s_t = j \mid \boldsymbol{\Omega}^t\big)$ is the best forecast of $s_t = j$ conditional on the information available at time t and also known as the filtering probability of $s_t = j$;

$$\zeta_{j,t-i} = \mathbb{P}\big(s_{t-i} = j \mid \boldsymbol{\Omega}^t\big) - \mathbb{P}\big(s_{t-i} = j \mid \boldsymbol{\Omega}^{t-1}\big)$$

is the revision of the perceived likelihoods (expectations) of the income innovation υ_{t-i} being permanent ($j = 1$) or transitory ($j = 0$) when the information set is expanded from $\boldsymbol{\Omega}^{t-1}$ to $\boldsymbol{\Omega}^t$. Note that $\zeta_{1,t-i} = -\zeta_{0,t-i}$, so that an upward revision on the perceived likelihood of state 1 must accompany with a downward revision on the perceived likelihood of state 0, and vice versa.

Equation (5) shows how a rational agent may employ available information to form the best forecast on the change in the permanent income (and hence the change in consumption). The first two terms on the right-hand side (RHS) of (5) characterizes the change of consumption due to the current innovation υ_t. As discussed in Deaton (1992), if labor income is modeled as $\Gamma(B)\Delta y_t = \alpha_0 + \upsilon_t$ so that all innovations are (known to be) permanent, such a change would be $\Gamma^* \upsilon_t$. When labor income is $\Psi(B) y_t = \upsilon_t$ with only transitory innovations, such a change becomes $\big[r/(1+r)\big] \Psi^* \upsilon_t$. In the present context, the agent faces uncertainty about the current innovation state and must draw the optimal inference $\mathbb{P}(s_t = j \mid \boldsymbol{\Omega}^t)$ for $j = 0, 1$. The change of consumption due to υ_t is thus a weighted average of the two distinct effects mentioned above, with the respective weights $\mathbb{P}(s_t = j \mid \boldsymbol{\Omega}^t)$. As these probabilities change over time, the effect of υ_t is also time varying.[2] Moreover, the change of consumption may also be influenced by past income

[2]We may also interpret $\mathbb{P}(s_t = 1 \mid \boldsymbol{\Omega}^t)\upsilon_t$ as a "permanent component" and $\mathbb{P}(s_t = 0 \mid \boldsymbol{\Omega}^t)\upsilon_t$ a "transitory component" of υ_t. Given that the consumption multipliers for a permanent innovation in income and a transitory one are, respectively, Γ^* and $[r/(1+r)]\Psi^*$, the first two terms on the RHS of equation (5) are the sum of the consumption changes resulted from these two components. Comparing with Quah (1990) and Falk and Lee (1998), the decomposition here hinges upon the conditional probabilities $\mathbb{P}(s_t = j \mid \boldsymbol{\Omega}^t)$ and hence may change over time.

innovations. As shown in the last two terms on the RHS of equation (5), the effects of past innovations depend on whether there is revision on the perceived likelihoods of their respective states. This is, again, a consequence of state uncertainty.

From the discussion above we can see that equation (5) is in sharp contrast with the results of Flavin (1981), Deaton (1987), Diebold and Rudebusch (1991) and others. First, the reaction of consumption to current innovation is a weighted average of two distinct effects and may vary with time. Second, consumption reacts not only to current innovation but also to past innovations (when there is revision on the perceived likelihoods of past states). In the existing models, the state of all innovations is known to the agent and never changes. The perceived likelihood of the known state is thus always one, while that of the other state is always zero. Because of this construction, a current innovations can have only a specific, time-invariant effect on the change of consumption, but past innovations can not have any effect since the perceived likelihoods can never be revised. In other words, the agent's consumption behavior would not be optimal if there is state uncertainty.

To illustrate, we again consider the simple case that the labor income follows a CD(1,1;1) process. It is readily verified that (5) now becomes

$$\begin{aligned}\Delta c_t = {} & \Gamma_1^* \mathbb{P}\big(s_t = 1 \mid \mathbf{\Omega}^t\big) v_t + \frac{r}{1+r} \Psi_1^* \mathbb{P}\big(s_t = 0 \mid \mathbf{\Omega}^t\big) v_t \\ & + \Gamma_1^* \sum_{i=1}^{\infty} \left(1 + \frac{r}{1+r} \sum_{j=1}^{i} \gamma_1^j\right) \zeta_{1,t-i} v_{t-i} + \frac{r}{1+r} \Psi_1^* \sum_{i=1}^{\infty} \psi_1^i \zeta_{0,t-i} v_{t-i},\end{aligned} \tag{6}$$

where $\Gamma_1^* = (1+r)/(1+r-\gamma_1)$ and $\Psi_1^* = (1+r)/(1+r-\psi_1)$. When the agent knows with certainty that the effects of innovations on future income are all permanent (i.e., $s_t = 1$ for all t), the perceived likelihoods of s_t are not affected by the information set and always equal to one. Hence, $\zeta_{1,t-i} = -\zeta_{0,t-i} = 0$ for all $i = 1, 2, \ldots$, so that Δc_t reacts only to the current innovation. In fact, $\Delta c_t = \Gamma_1^* v_t$ for all t, which is exactly the result obtained in Deaton (1987) for an ARIMA(1,1,0) income process. Similarly, when the agent knows a priori that the effects of income innovations are all transitory (i.e., $s_t = 0$ for all t), Δc_t is still determined by the current innovation: $\Delta c_t = [r/(1+r)]\Psi_1^* v_t$ for all t. This is the result of Flavin (1981) based on an AR(1) income process. This example shows that, when there is no state uncertainty, existing results are in fact special cases of ours. The proposed CD model thus provides a much more general framework for analyzing the consumption behavior.

Moreover, the marginal propensity to consumption (MPC) due to an unanticipated change in labor income is:

$$\mathrm{MPC} \equiv \frac{\Delta c_t}{\upsilon_t}\bigg|_{\{\upsilon_{t-i}=0\}} = \Gamma^* \,\mathbb{P}\big(s_t = 1 \mid \mathbf{\Omega}^t\big) + \frac{r}{1+r}\Psi^* \,\mathbb{P}\big(s_t = 0 \mid \mathbf{\Omega}^t\big).$$

This MPC depends on the perceived likelihood of the innovation state and hence may also change over time. By contrast, the MPCs obtained in Flavin (1981), Deaton (1987) and others are time invariant. Given the realization of an innovation state, the MPCs for permanent and transitory income innovations are, respectively,

$$\mathrm{MPC}_1 \equiv \mathrm{MPC}_{s_t=1} = \Gamma^*,$$

$$\mathrm{MPC}_0 \equiv \mathrm{MPC}_{s_t=0} = \frac{r}{1+r}\Psi^*.$$

We may empirically examine an important implication of the PIH based on these MPCs. That is, we may test whether MPC_1 is significantly greater than MPC_0, cf. Hall and Mishkin (1982).

4 Model Estimation and Hypothesis Testing

To estimate the income process (4) and consumption formula (5), we first postulate that s_t follows a two-state Markov chain with the transition matrix

$$\begin{bmatrix} \mathbb{P}\big(s_t = 0 \mid s_{t-1} = 0\big) & \mathbb{P}\big(s_t = 1 \mid s_{t-1} = 0\big) \\ \mathbb{P}\big(s_t = 0 \mid s_{t-1} = 1\big) & \mathbb{P}\big(s_t = 1 \mid s_{t-1} = 1\big) \end{bmatrix} = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix},$$

where $p_{i0} + p_{i1} = 1$ for $i = \{0, 1\}$. It has been emphasized in the literature that the individual's own consumption behavior should provide a good instrument for revealing his/her expectations about future labor income; see e.g., Campbell and Deaton (1989), Campbell and Mankiw (1990) and Flavin (1993). This suggests us to estimate (4) and (5) jointly. To render estimation tractable, we employ a finite number of terms to approximate the infinite order of revisions on the expectations about $\{s_{t-i}\}$ in (5) and use a random term to summarize the errors arising from this approximation.

In our empirical study, the following system of equations is estimated:

$$\begin{aligned} \Gamma(B)\Psi(B)\Delta y_t &= \alpha_0 \Psi(1) + \sum_{i=1}^{\kappa+1} \xi_{i,s_{t-i}} \upsilon_{t-i} + \upsilon_t, \\ \Delta c_t &= \varphi_1 \,\mathbb{P}\big(s_t = 1 \mid \mathbf{\Omega}^t\big)\upsilon_t + \varphi_0 \,\mathbb{P}\big(s_t = 0 \mid \mathbf{\Omega}^t\big)\upsilon_t + \phi_1 \zeta_{1,t-1}\upsilon_{t-1} + \varepsilon_t, \end{aligned} \tag{7}$$

where φ_1, φ_0, and ϕ_1 are parameters, ε_t are random errors with mean zero and variance σ_ε^2. Note that following Hall and Mishkin (1982), the responses of consumption to the unanticipated changes components are estimated as free parameters (i.e., φ_1 and φ_0). Under weak assumptions, it can be shown that υ_t and ε_t are uncorrelated at all leads and lags.

The system of equations (7) contains the parameters:

$$\boldsymbol{\theta} = (\gamma_1, \ldots, \gamma_n, \psi_1, \ldots, \psi_m, \alpha_0, \varphi_0, \varphi_1, \phi_1, \sigma_\upsilon^2, \sigma_\varepsilon^2, p_{00}, p_{11})',$$

which can be estimated by either the (approximate) quasi-maximum likelihood method or Markov chain Monte Carlo method. We adopt the former in this paper and follow the agent's decision rule discussed in the previous section to construct the estimation algorithm. In each period, an innovation υ_t to the labor income occurs, and the agent obtains the labor earnings y_t. With these information, the agent revises his/her expectations about the states of the past innovations via the Bayes' rule. The agent then forms his/her expectations about the state of the current income innovation and determines the amount of money to spend. A detailed derivation of the estimation algorithm is given in Appendix II.

Since many studies have established that the labor income process may contain a unit root (e.g., Deaton, 1987, and West, 1988), it is imperative to test the labor income as a CD process versus an ARIMA process. In the present context, this amounts to testing whether $p_{11} = 1$. Under the null hypothesis, the transitory component does not enter the model so that the parameters in $\Psi(B)$ are not identified. In this case, standard likelihood-based tests are not applicable, as discussed in Davies (1977, 1987) and Hansen (1996). To circumvent this problem, we adopt the simulation-based test proposed by KHT. Specifically, we first estimate an array of ARIMA$(r, 1, q)$ models for labor income and choose the best specification based on an information criterion (e.g., AIC or SIC). The selected model will be denoted as ARIMA$(\tilde{r}^*, 1, \tilde{q}^*)$. We also estimate an array of CD$(n, 1; m)$ models in (4) for labor income and choose the best model based on AIC or SIC. We denote the selected model as CD$(\tilde{n}^*, 1; \tilde{m}^*)$ and the estimate of the transition probability as $\tilde{p}_{11}^*$. The selected ARIMA$(\tilde{r}^*, 1, \tilde{q}^*)$ model will be taken as the data generating process to generate simulated samples. For each simulated sample, we re-estimate the CD$(\tilde{n}^*, 1; \tilde{m}^*)$ model and obtain an estimate of p_{11}, denoted as $\tilde{p}_{11}$. Replicating this procedure many times yields an empirical distribution of $\tilde{p}_{11}$. We then compare $\tilde{p}^*_{11}$ with the quantiles of this empirical distribution. The null hypothesis that the labor income follows the ARIMA$(\tilde{r}^*, 1, \tilde{q}^*)$ model would be rejected if the empirical p-value of $\tilde{p}^*_{11}$ is less than, say, 5%.

Once the test above suggests that the labor income follows a CD$(n, 1; m)$ process, we can proceed to test the implications of the PIH. There are two types of tests. One focuses on the validity of the restrictions on the parameters implied by the PIH. This is done by comparing the unconstrained parameter estimates of φ_1, φ_0, and ϕ_1 in (7) with the corresponding annuity values predicted by the PIH (see Section 5.2 for explicit expressions). In addition, we also examine an important implication of the PIH: a permanent income innovation exerts a much larger effect on consumption than does a transitory one. This amounts to testing $\varphi_1 = \varphi_0$ against $\varphi_1 > \varphi_0$. This hypothesis is tested by checking whether the difference between their estimates is significantly greater than zero.

5 Empirical Analysis

To assess the empirical relevance of the proposed model, we estimate (7) based on U.S. consumption and income data. The data are real personal disposable labor income per capita and real consumption (non-durable goods and services) per capita from 1959:III through 1999:II, a total of 160 observations. These series are constructed using the Blinder and Deaton (1985) procedures and taken as the variables y_t and c_t in model (7), respectively.

5.1 Model Estimation Results

We first apply the simulation-based test discussed in the preceding section to check whether the labor income is actually an ARIMA process. We estimate an array of ARIMA$(r, 1, q)$ models for y_t with r and q no greater than 3; the best model based on AIC and SIC select is the following ARIMA(1,1,0) model:

$$\Delta y_t = 12.416 + 0.086 \Delta y_{t-1} + e_t, \tag{8}$$

with $\sigma_e = 27.898$. We also estimate an array of CD$(n, 1; m)$ models for y_t with m and n no greater than 3; both AIC and SIC select the CD(1,1;1) model. For this CD model, the estimated transition probabilities are $\tilde{p}^*_{11} = 0.896$ and $\tilde{p}^*_{00} = 0.782$.[3] For each simulated

[3]The estimation results of these ARIMA and CD models are available from the authors upon request.

Table 2: Quasi-maximum likelihood estimates of the proposed model in (7).

Parameter	Estimate	Standard error	t-statistic
γ_1	0.307	0.084	16.271*
ψ_1	0.204	0.123	10.147*
α_0	0.227	3.534	0.064
φ_0	0.049	0.138	0.357
φ_1	0.511	0.076	6.675*
ϕ_1	0.433	0.217	1.994*
σ_υ	25.068	2.319	10.803*
σ_ε	14.867	0.950	15.649*
p_{00}	0.773	0.088	
p_{11}	0.876	0.046	
Log-Likelihood=-1416.08		AIC=2852.169	SIC=2882.857

Note: t-statistics with an asterisk are significant at 5% level.

sample generated according to equation (8), we re-estimate the CD(1,1;1) model to get an estimate of p_{11}. With 1000 replications, we obtain an empirical distribution of $\tilde{p}_{11}$; the empirical p-value of $\tilde{p}_{11}^* = 0.896$ is 0.041. Therefore, we are able to reject the hypothesis that the data are generated from (8) at 5% level.

With the testing result above, we proceed to estimate the system in (7) with m and n no greater than 3. The parameters $\boldsymbol{\theta}$ here are estimated using the algorithm described in Appendix II. This algorithm is initialized by a broad range of random initial values. The covariance matrix of $\boldsymbol{\theta}$ is $-\boldsymbol{H}(\hat{\boldsymbol{\theta}}_T)^{-1}$, the Hessian matrix of the log-likelihood function evaluated at the QMLE $\hat{\boldsymbol{\theta}}_T$. Among all the models estimated, both AIC and SIC select the one with $m = n = 1$. The estimation results are summarized in Table 2; in particular, the estimated transition probabilities ($\hat{p}_{11} = 0.876$ and $\hat{p}_{00} = 0.773$) are quite close to those obtained from estimating the CD(1,1;1) model alone.

We also conduct some diagnostic checks on the estimated model, including the Ljung-Box (1978) Q test and the LM test of Engle (1982) on the ARCH effect. The resulting statistics for the income residuals $\hat{\upsilon}_t$ are $Q(20) = 23.378$ and ARCH(4) $= 5.736$, and those for the consumption residuals $\hat{\varepsilon}_t$ are $Q(20) = 15.435$ and ARCH(4) $= 2.855$. These statistics are all insignificant even at 10% level under the $\chi^2(20)$ and $\chi^2(4)$ distributions.

Hence, there appears no serial correlation and conditional heteroskedasticity in these residuals. Following Engel and Hamilton (1990), we test whether the state variables are independent over time, i.e., $p_{00} + p_{11} = 1$. The resulting Wald statistic is 42.797 and rejects the null at 1% level under the $\chi^2(1)$ distribution. This result is consistent with the Markovian specification.

In Section 3 we emphasize that effect of a current innovation on the change of consumption hinges on the filtering probabilities $\mathbb{P}(s_t = j \mid \boldsymbol{\Omega}^t; \boldsymbol{\theta})$, $j = 1, 0$. In Figure 1, we plot $\mathbb{P}(s_t = 1 \mid \boldsymbol{\Omega}^t; \hat{\boldsymbol{\theta}}_T)$, the filtering probabilities of $s_t = 1$ evaluated at the parameter estimate $\hat{\boldsymbol{\theta}}_T$. As these probabilities vary substantially over time, so do the effects of current innovations on consumption. It can also be seen that there are 21 periods (about 14% of the sample) in which $\mathbb{P}(s_t = 1 \mid \boldsymbol{\Omega}^t; \hat{\boldsymbol{\theta}}_T) < 0.5$. That is, more than 80 percent of income innovations are more likely to be treated as a permanent innovation at the time they occur. However, the ergodic probability of $s_t = 1$ is

$$\mathbb{P}(s_t = 1) \equiv \lim_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\sum_{t=0}^{T-1} \mathbf{1}_{\{s_t=1\}}\right] = \frac{1 - p_{00}}{2 - p_{00} - p_{11}} \approx 65\%,$$

where $\mathbf{1}$ is the indicator function of $s_t = 1$. This suggests that approximately 65 percent of the innovations may have a permanent effect in the long-run. Both results indicate that not all labor income innovations have a permanent effect, contrary to the specification of Deaton (1987) and West (1988). Similarly, not all income innovations are transitory, cf. Flavin (1981).

In Section 3 we also stress the importance of the revision of the expectations on past states. In the estimation result of (7), $\hat{\phi}_1$ is significantly different from zero, suggesting a significant effect of such revision on the change of consumption. In Figure 2, we plot

$$\hat{\zeta}_{1,t} = \mathbb{P}(s_t = 1 \mid \boldsymbol{\Omega}^{t+1}; \hat{\boldsymbol{\theta}}_T) - \mathbb{P}(s_t = 1 \mid \boldsymbol{\Omega}^t; \hat{\boldsymbol{\theta}}_T),$$

and find that there are more positive (upward) revisions than negative (downward) ones, where the former indicates reinforcement of the perceived likelihood when the information set enlarges. Moreover, the range of these revisions is quite substantial. In particular, $|\hat{\zeta}_{1,t}|$ has the sample mean (sample standard deviation) of is 0.122 (0.131) and the maximum 0.725. Consequently, neglecting these revisions may result in very misleading consumption decisions.

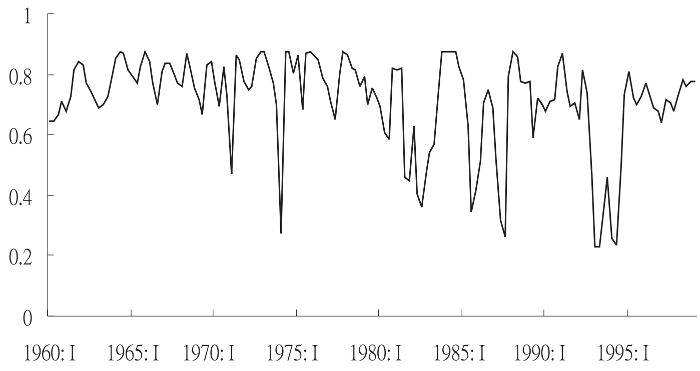

Figure 1: Estimated filtering probabilities of $s_t = 1$.

5.2 Testing the Implications of the PIH

We now proceed to test the cross-equations restrictions in (7) as implied by the PIH and the CD labor income process. These restrictions include:

$$\varphi_1 = \Gamma^* = \frac{1+r}{1+r-\gamma_1},$$

$$\varphi_0 = \frac{r}{1+r}\Psi^* = \frac{r}{1+r-\psi_1},$$

$$\phi_1 = \Gamma^*\left[1 + \frac{r}{1+r}\gamma_1\right] - \frac{r}{1+r}\Psi^*\psi_1,$$

where φ_1, φ_0 and ϕ_1 are the parameters of (7), and the second equality on the RHS of each equation above represents the prediction of the PIH. We assume that the real interest rate is a constant 0.01 (i.e., 4% per annum).

For the first restriction $\varphi_1 = \Gamma^*$, the unconstrained estimate of MPC_1 is $\hat{\varphi}_1 = 0.511$, and given our estimate of γ_1 is 0.307, the PIH implied MPC_1 is $\hat{\Gamma}^* = 1.436$. The Wald test is , which is significant at 5% level. Thus, the actual consumption under-reacts to permanent innovations in labor income when compared with the prediction of the PIH. For the second restriction $\varphi_0 = \Psi^* r/(1+r)$, the unconstrained estimate of MPC_0 is $\hat{\varphi}_0 = 0.049$, the PIH implied MPC_0 is $\hat{\Psi}^* r/(1+r) = 0.010$, and the Wald statistic is 0.069, which is insignificant at 5% level. This shows that consumption does not appear to over-react to a transitory innovation in labor income. For the last restriction, the unconstrained estimate is $\hat{\phi}_1 = 0.433$, and that implied by the PIH is 1.438. The Wald statistic is 0.091, which is significant at the 5% level.

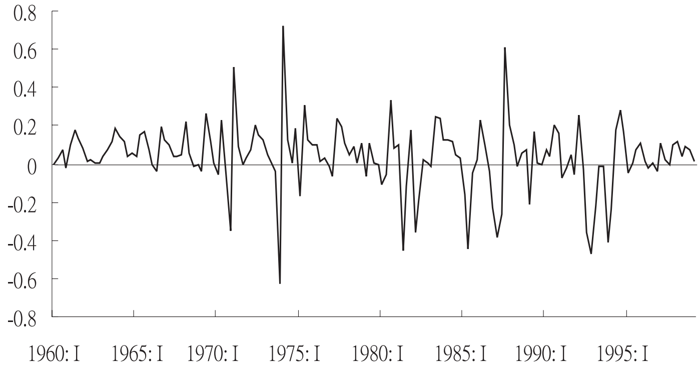

Figure 2: The revision in the estimated filtered probability of the permanent state: $\hat{\zeta}_{1,t}$.

Even though some of our estimation results conflict with the prediction of the PIH, they are not entirely against the intertemporal consumption model. In particular, the estimated response of consumption to a permanent innovation in income, ($\hat{\varphi}_1$) is much greater than that to a transitory innovation ($\hat{\varphi}_0$). The Wald statistic of the null hypothesis $\varphi_1 = \varphi_0$ is 6.688 which is significant even at the 1% level under the $\chi^2(1)$ distribution. This result reveals certain evidence for the agent to allocate his/her labor income intertemporally in determining the current consumption, even though the consumption decision here does not entirely follow the prediction of the PIH.

In Figure 3, we plot two MPC series constructed from our estimation results, where the dashed line denotes the MPC implied by the PIH and the estimated CD model of labor income:

$$\Gamma^* \,\mathbb{P}\big(s_t = 1 \mid \mathbf{\Omega}^t\big) + \frac{r}{1+r}\Psi^* \,\mathbb{P}\big(s_t = 0 \mid \mathbf{\Omega}^t\big),$$

and the solid line is the MPC derived from unconstrained estimation of (7):

$$\varphi_1 \,\mathbb{P}\big(s_t = 1 \mid \mathbf{\Omega}^t\big) + \varphi_0 \,\mathbb{P}\big(s_t = 0 \mid \mathbf{\Omega}^t\big).$$

The figure shows that the solid line is generally smaller and smoother than the dashed line. This difference is mainly due to the fact that $\hat{\varphi}_1$ is significantly smaller than that implied by the PIH.

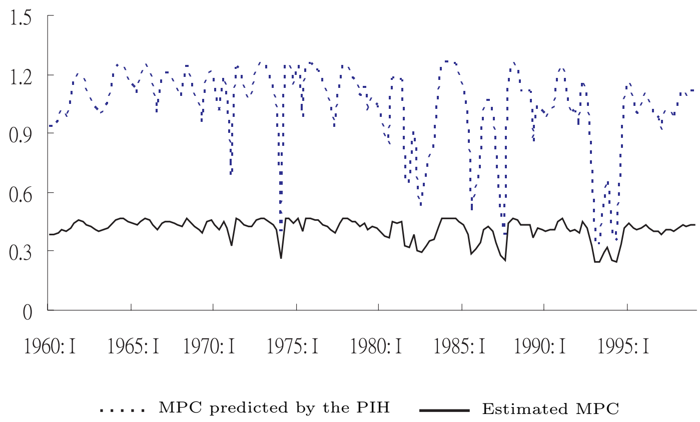

Figure 3: The time-varying MPC for U.S. real consumption.

6 Conclusion

In this paper we take a different modeling approach based on the CD model of KHT to testing the PIH, an example of intertemporal choice models. A key feature of this approach is that it explicitly allows for state uncertainty. In the existing literature of intertemporal models, the state of innovations is assumed to be fixed and known to the agent a priori. This somewhat simplistic assumption may not hold in the real world. When state uncertainty is present, the optimal decision of the agent ought to depend on his/her forecasts on the innovation state. Unfortunately, existing models have no room for drawing such forecasts, nor do they allow the agent to adjust his/her behavior based on available information. The resulting behavior thus can not be truly optimal under state uncertainty. By taking state uncertainty into account, our modeling approach provides a more general framework to analyze a rational agent's intertemporal decisions.

In this paper, it is shown that when the labor income follows a CD process with unobservable innovation state, the consumption behavior predicted by the PIH is more complicated than those obtained from traditional models, such as Flavin (1981) and Deaton (1987). In particular, the reaction of consumption to the current innovation in labor income depends on the agent's forecast on the innovation state. Moreover, consumption is affected by past innovations when there is revision in the expectations about past innovation states. To our knowledge, the effect of past innovations has not been documented in the literature. The resulting MPC due to an unanticipated change in labor income also depends on how the agent forecasts the innovation state and hence is time varying. Without state uncertainty, the consumption behavior derived in this paper simply reduces to that obtained in the literature.

Our empirical study shows that the proposed model characterizes U.S. real consumption and income data very well. Although the estimation results do not always agree with the predictions of the PIH, they are not entirely against the intertemporal consumption model. It should be emphasized that the PIH per se is not the major concern of this paper. What we try to demonstrate is the importance of considering state uncertainty in intertemporal choice models. Since the identification of the innovation state of some variables is usually a crucial step for understanding the strength and weakness of the intertemporal rational expectations model studied, the CD modeling approach proposed in the paper should serve as a useful and powerful tool for future studies on the related topics as such.

Appendix I

Let $\mathcal{S}^t = \{s_t, s_{t-1}, \ldots\}$ denote the collection of all current and past state variables, and $\tilde{\boldsymbol{\Omega}}^t = \mathcal{S}^t \cup \boldsymbol{\Omega}^t$ be the full information set about income innovations up to time t. We also assume that $\{\upsilon_t\}$ is a sequence of random variables such that $\mathbb{E}(\upsilon_t \mid \mathcal{S}^t, \boldsymbol{\Omega}^{t-1}) = 0$ and $\operatorname{var}(\upsilon_t \mid \mathcal{S}^t, \boldsymbol{\Omega}^{t-1}) = \sigma_\upsilon^2$. By invoking the law of iterated expectations, it is easy to verify that $\{\upsilon_t\}$ is a white noise and $\mathbb{E}(s_t \upsilon_t \mid \boldsymbol{\Omega}^{t-1}) = 0$. Also, $\mathbb{E}(s_t \upsilon_t) = 0$,

$$\operatorname{var}(s_t \upsilon_t) = \mathbb{E}\big[s_t^2\, \mathbb{E}(\upsilon_t^2 \mid \mathcal{S}^t, \boldsymbol{\Omega}^{t-1})\big] = \sigma_\upsilon^2\, \mathbb{P}(s_t = 1),$$

and $\operatorname{cov}(s_t \upsilon_t, s_{t-i} \upsilon_{t-i}) = \mathbb{E}\big[s_t s_{t-i} \upsilon_{t-i}\, \mathbb{E}(\upsilon_t \mid \mathcal{S}^t, \boldsymbol{\Omega}^{t-1})\big] = 0$. Similarly, $(1 - s_t)\upsilon_t$ has mean zero and variance $\big[1 - \mathbb{P}(s_t = 1)\big]\sigma_\upsilon^2$ and are serially uncorrelated. Hence, these two series are white noise when $\mathbb{P}(s_t = 1)$ is a constant π_0.

Under the assumptions above, the components $\Delta y_{1,t}$ and $y_{0,t}$ in (3) can be viewed as stationary ARMA processes with serially uncorrelated innovations $s_t \upsilon_t$ and $(1 - s_t)\upsilon_t$, respectively. Directly applying the Wiener-Kolmogorov prediction formula discussed in Hansen and Sargent (1980, p. 16), we obtain explicit solutions to the prediction problems for $y_{1,t+i}$ and $y_{0,t+i}$, conditional on the full information set $\tilde{\boldsymbol{\Omega}}^t$:

$$\begin{aligned}
\sum_{i=0}^{\infty} (1+r)^{-i}\, \tilde{\mathbb{E}}_t y_{1,t+i} &= \frac{1+r}{r} \Gamma^* \left[1 + \sum_{j=1}^{n} \Big(\sum_{k=j+1}^{n+1} \big(\frac{1}{1+r}\big)^{k-j} \tilde{\gamma}_k \Big) B^j \right] y_{1,t}, \\
\sum_{i=0}^{\infty} (1+r)^{-i}\, \tilde{\mathbb{E}}_t y_{0,t+i} &= \Psi^* \left[1 + \sum_{j=1}^{m-1} \Big(\sum_{k=j+1}^{m} \big(\frac{1}{1+r}\big)^{k-j} \psi_k \Big) B^j \right] y_{0,t},
\end{aligned} \tag{9}$$

where $\tilde{\mathbb{E}}_t = \mathbb{E}(\,\cdot \mid \tilde{\boldsymbol{\Omega}}^t)$ denotes the expectation conditional on $\tilde{\boldsymbol{\Omega}}^t$, the values Γ^* and Ψ^* are the polynomials $\Gamma^{-1}(B)$ and $\Psi^{-1}(B)$ evaluated at the discount factor $1/(1+r)$, and $\tilde{\gamma}_k$ are the coefficients of $(1-B)\Gamma(B) = 1 - \tilde{\gamma}_1 B - \cdots - \tilde{\gamma}_{n+1} B^{n+1}$. Substituting (3) into (9) yields

$$\begin{aligned}
\sum_{i=0}^{\infty} (1+r)^{-i}\, \tilde{\mathbb{E}}_t y_{1,t+i} &= \frac{1+r}{r} \Gamma^* \sum_{i=0}^{\infty} \gamma_i^* s_{t-i} \upsilon_{t-i}, \\
\sum_{i=0}^{\infty} (1+r)^{-i}\, \tilde{\mathbb{E}}_t y_{0,t+i} &= \Psi^* \sum_{i=0}^{\infty} \psi_i^* (1 - s_{t-i}) \upsilon_{t-i},
\end{aligned} \tag{10}$$

where the terms γ_i^* and ψ_i^* are the coefficients of

$$\left[1 + \sum_{j=1}^{n} \Big(\sum_{k=j+1}^{n+1} \big(\frac{1}{1+r}\big)^{k-j} \tilde{\gamma}_k \Big) B^j \right] \Gamma(B)^{-1}$$

and

$$\left[1+\sum_{j=1}^{m-1}\Bigg(\sum_{k=j+1}^{m}\big(\frac{1}{1+r}\big)^{k-j}\psi_k\Bigg)B^j\right]\Psi(B)^{-1},$$

respectively. By the law of iterated expectations, we thus have

$$\sum_{i=0}^{\infty}(1+r)^{-i}\,\mathbb{E}_t y_{j,t+i} = \mathbb{E}_t\Bigg(\sum_{i=0}^{\infty}(1+r)^{-i}\,\tilde{\mathbb{E}}_t y_{j,t+i}\Bigg),$$

$$\sum_{i=0}^{\infty}(1+r)^{-i}\,\mathbb{E}_{t-1} y_{j,t+i} = \mathbb{E}_{t-1}\Bigg(\sum_{i=0}^{\infty}(1+r)^{-i}\,\mathbb{E}_t y_{j,t+i}\Bigg),$$

for $j = 0, 1$. Substituting these equations into the consumption formula in (2) and rearranging, we obtain (5):

$$\begin{aligned}\Delta c_t &= \frac{r}{1+r}\sum_{i=0}^{\infty}(1+r)^{-i}\big(\mathbb{E}_t - \mathbb{E}_{t-1}\big)\big(y_{0,t+i}+y_{1,t+i}\big)\\ &= \Gamma^*\,\mathbb{P}\big(s_t = 1 \mid \boldsymbol{\Omega}^t\big)\upsilon_t + \frac{r}{1+r}\Psi^*\,\mathbb{P}\big(s_t = 0 \mid \boldsymbol{\Omega}^t\big)\upsilon_t\\ &\quad + \Gamma^*\sum_{i=1}^{\infty}\gamma_i^*\zeta_{1,t-i}\upsilon_{t-i} + \frac{r}{1+r}\Psi^*\sum_{i=1}^{\infty}\psi_i^*\zeta_{0,t-i}\upsilon_{t-i}.\end{aligned}$$

Appendix II

To simplify the exposition, we first define some notations needed in this Appendix. From equation (7) we see that the past $\kappa+1$ state variables affect Δy_t. Following Hamilton (1994), we define the new state variable $s^*_{t-1} = 1, 2, \ldots, 2^{\kappa+1}$ such that each of these values represents a particular combination of the realizations of $(s_{t-1}, \ldots, s_{t-\kappa-1})$. It is easy to show that s^*_t also forms a first-order Markov chain with the transition matrix $\boldsymbol{P}^*$. This transition matrix can be expressed as

$$\boldsymbol{P}^* = \begin{bmatrix} \boldsymbol{P}_{00} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{P}_{10} \\ \boldsymbol{P}_{01} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{P}_{11} \end{bmatrix},$$

with $\boldsymbol{P}_{ji}$ $(j, i = 0, 1)$ being a $2^{\kappa-1} \times 2^{\kappa}$ block diagonal matrix given by

$$\boldsymbol{P}_{ji} = \begin{bmatrix} p_{ji} & p_{ji} & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & p_{ji} & p_{ji} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & p_{ji} & p_{ji} \end{bmatrix}.$$

Also let $\boldsymbol{\upsilon}_{t-1} = (\upsilon_{t-1}, \ldots, \upsilon_{t-\kappa-1})'$ and for $s^*_{t-1} = \ell$, $\ell = 1, 2, \ldots, 2^{\kappa+1}$, let

$$\boldsymbol{\xi}_{t-1,\ell} = (\xi_{1,s_{t-1}}, \xi_{2,s_{t-2}}, \ldots, \xi_{\kappa+1,s_{t-\kappa-1}})',$$

where the realizations of $s_{t-1}, \ldots, s_{t-\kappa-1}$ are such that $s^*_{t-1} = \ell$. Then,

$$\boldsymbol{\xi}'_{t-1,\ell}\boldsymbol{\upsilon}_{t-1} = \sum_{j=1}^{\kappa+1} \xi_{j,s_{t-j}} \upsilon_{t-j}$$

in equation (7).

To derive the estimation algorithm, we first discuss the optimal forecasts of the state variable s_t based on the information up to time t. Under the normality assumption, the density of Δy_t conditional on $s^*_{t-1} = \ell$ and $\boldsymbol{\Omega}^{t-1}$ is

$$\begin{aligned} & f\big(\Delta y_t \mid s^*_{t-1} = \ell, \boldsymbol{\Omega}^{t-1}; \boldsymbol{\theta}\big) \\ & \quad = \frac{1}{\sqrt{2\pi\sigma^2_\upsilon}} \exp\left\{ \frac{-\big[\Gamma(B)\Psi(B)\Delta y_t - \alpha_0\Psi(1) - \boldsymbol{\xi}'_{t-1,\ell}\boldsymbol{\upsilon}_{t-1}\big]^2}{2\sigma^2_\upsilon} \right\}, \end{aligned} \tag{11}$$

where $\ell = 1, 2, \cdots, 2^{\kappa+1}$ and

$$\boldsymbol{\theta} = (\gamma_1, \ldots, \gamma_n, \psi_1, \ldots, \psi_m, \alpha_0, \varphi_0, \varphi_1, \phi_1, \sigma^2_\upsilon, \sigma^2_\varepsilon, p_{00}, p_{11})'.$$

Although the innovations υ_t depend on s^*_{t-1} $(t = m+1, \ldots, T)$, we follow Gray (1996) and compute υ_t $(t = m+1, \ldots, T)$ as

$$\begin{aligned} \upsilon_t &= \Delta y_t - \mathbb{E}(\Delta y_t \mid \boldsymbol{\Omega}^{t-1}) \\ &= \Gamma(B)\Psi(B)\Delta y_t - \alpha_0\Psi(1) - \sum_{\ell=1}^{2^{\kappa+1}} \mathbb{P}\big(s^*_{t-1} = \ell \mid \boldsymbol{\Omega}^{t-1}; \boldsymbol{\theta}\big)\, \boldsymbol{\xi}'_{t-1,\ell}\boldsymbol{\upsilon}_{t-1}, \end{aligned} \tag{12}$$

with the initial values $\upsilon_m, \ldots, \upsilon_1$ being zero, where $\mathbb{P}\big(s^*_{t-1} = \ell \mid \boldsymbol{\Omega}^{t-1}; \boldsymbol{\theta}\big)$ is the probability of $s^*_{t-1} = \ell$ based on the information up to time $t-1$.

Given the optimal forecasts of the previous state variables $\mathbb{P}\big(s_{t-1}^* = \ell \mid \boldsymbol{\Omega}^{t-1}; \boldsymbol{\theta}\big)$, the density of Δy_t conditional on $\boldsymbol{\Omega}^{t-1}$ alone can be obtained via (11) as

$$f\big(\Delta y_t \mid \boldsymbol{\Omega}^{t-1}; \boldsymbol{\theta}\big) = \sum_{\ell=1}^{2^{\kappa+1}} \mathbb{P}\big(s_{t-1}^* = \ell \mid \boldsymbol{\Omega}^{t-1}; \boldsymbol{\theta}\big)\, f\big(\Delta y_t \mid s_{t-1}^* = \ell, \boldsymbol{\Omega}^{t-1}; \boldsymbol{\theta}\big). \tag{13}$$

Based on the new information at time t, the expectations about the states of past innovations will be revised according to the Bayes's theorem:

$$\mathbb{P}\big(s_{t-1}^* = \ell \mid \boldsymbol{\Omega}^{t}; \boldsymbol{\theta}\big) = \frac{\mathbb{P}\big(s_{t-1}^* = \ell \mid \boldsymbol{\Omega}^{t-1}; \boldsymbol{\theta}\big)\, f\big(\Delta y_t \mid s_{t-1}^* = \ell, \boldsymbol{\Omega}^{t-1}; \boldsymbol{\theta}\big)}{f\big(\Delta y_t \mid \boldsymbol{\Omega}^{t-1}; \boldsymbol{\theta}\big)}. \tag{14}$$

Then, to form optimal forecasts about the current state variables $\mathbb{P}\big(s_t^* = \ell \mid \boldsymbol{\Omega}^{t}; \boldsymbol{\theta}\big)$ based on the new information, we assume that the (j,i)th element of $\boldsymbol{P}^*$ is such that

$$p_{ji}^* = \mathbb{P}\big(s_t^* = i \mid s_{t-1}^* = j\big) = \mathbb{P}\big(s_t^* = i \mid s_{t-1}^* = j, \boldsymbol{\Omega}^{t}\big);$$

the second equality would hold if $\{s_t\}$ and $\{\upsilon_t\}$ are independent. These in turn yield

$$\begin{aligned} \mathbb{P}\big(s_t^* = \ell \mid \boldsymbol{\Omega}^{t}; \boldsymbol{\theta}\big) &= \sum_{j=1}^{2^{\kappa+1}} \mathbb{P}\big(s_{t-1}^* = j \mid \boldsymbol{\Omega}^{t}; \boldsymbol{\theta}\big)\, \mathbb{P}\big(s_t^* = \ell \mid s_{t-1}^* = j, \boldsymbol{\Omega}^{t}; \boldsymbol{\theta}\big) \\ &= \sum_{j=1}^{2^{\kappa+1}} p_{j\ell}^*\, \mathbb{P}\big(s_{t-1}^* = j \mid \boldsymbol{\Omega}^{t}; \boldsymbol{\theta}\big). \end{aligned} \tag{15}$$

Given the assumptions described in Appendix I, it is easy but tedious to show that υ_t and ε_t are uncorrelated at all leads and lags. Hence, with the normality assumption, the density of Δc_t conditional on $\boldsymbol{\Omega}^t$ is

$$f\big(\Delta c_t \mid \boldsymbol{\Omega}^{t}; \boldsymbol{\theta}\big) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left\{ -\frac{[\Delta c_t - \mu_t]^2}{2\,\sigma_\varepsilon^2} \right\}, \tag{16}$$

where $\mu_t = \varphi_0\, \mathbb{P}\big(s_t = 0 \mid \boldsymbol{\Omega}^{t}; \boldsymbol{\theta}\big)\upsilon_t + \varphi_1\, \mathbb{P}\big(s_t = 1 \mid \boldsymbol{\Omega}^{t}; \boldsymbol{\theta}\big)\upsilon_t + \phi_1 \zeta_{1,t-1}\upsilon_{t-1}$. Given the equations (12), (14), (15) and the initial probabilities $\mathbb{P}\big(s_{t-1}^* = \ell \mid \boldsymbol{\Omega}^{t-1}; \boldsymbol{\theta}\big)$, we can calculate the conditional density of Δc_t in (16). For example, the filtering probability $\mathbb{P}(s_t = 1 \mid \boldsymbol{\Omega}^t)$ and the revision $\zeta_{1,t-1}$ can be obtained via $\sum \mathbb{P}\big(s_t^* = \ell \mid \boldsymbol{\Omega}^{t}; \boldsymbol{\theta}\big)$ and

$$\sum \mathbb{P}\big(s_{t-1}^* = \ell \mid \boldsymbol{\Omega}^{t}; \boldsymbol{\theta}\big) - \sum \mathbb{P}\big(s_{t-1}^* = \ell \mid \boldsymbol{\Omega}^{t-1}; \boldsymbol{\theta}\big),$$

where the summation is taken over all ℓ that associated with $s_{t-i} = 1$ for $i = 0, 1$.

With the initial value $\mathbb{P}\big(s_m^* \mid \boldsymbol{\Omega}^m;\boldsymbol{\theta}\big)$, we can iterate the equations (11)–(16) to obtain $\mathbb{P}\big(s_t^* = \ell \mid \boldsymbol{\Omega}^t;\boldsymbol{\theta}\big)$ for $t = m+1,\ldots,T$. From the recursions above we also obtain the quasi-log-likelihood function

$$\log \mathcal{L}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \log f\big(\Delta y_t \mid \boldsymbol{\Omega}^{t-1};\boldsymbol{\theta}\big) + \log f\big(\Delta c_t \mid \boldsymbol{\Omega}^{t};\boldsymbol{\theta}\big),$$

from which the quasi-maximum likelihood estimator $\hat{\boldsymbol{\theta}}_T$ can be computed using a numerical algorithm. The estimation program is written in GAUSS which employs the BFGS (Broyden-Fletcher-Goldfarb-Shanno) algorithm. Following Hamilton (1989, 1994), we set the initial value $\mathbb{P}\big(s_m^* \mid \boldsymbol{\Omega}^m;\boldsymbol{\theta}\big)$ to its limiting unconditional counterpart: the $(2^{\kappa+1}+1)$th column of the matrix $(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$, where

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{I} - \boldsymbol{P}^* \\ \boldsymbol{1}' \end{bmatrix},$$

$\boldsymbol{I}$ is the identity matrix and $\boldsymbol{1}$ is the $2^{\kappa+1}$-dimensional vector of ones; see Hamilton (1994, p. 684) for details.

Reference

Abel, A. B. and O. Blanchard (1986). The present value of profits and cyclical movements in investment, *Econometrica*, **51**, 675–692.

Ahmed, S. (1986). Temporary and permanent government spending in an open economy, *Journal of Monetary Economics*, **17**, 197–224.

Barro, R. J. (1979). On the determination of the public debt, *Journal of Political Economy*, **82**, 1095–1117.

Barro, R. J. (1981). Output effects of government purchases, *Journal of Political Economy*, **89**, 1086–1121.

Beaudry, P. and G. Koop (1993). Do recessions permanently change output? *Journal of Monetary Economics*, **31**, 149–163.

Blinder, A. S. and A. Deaton (1985). The time series consumption function revisited, *Brookings Papers on Economic Activity*, **2**, 465–521.

Bradley, M. D. and D. W. Jansen (1997). Nonlinear business cycle dynamics: Cross-country evidence on the persistence of aggregate shocks, *Economic Inquiry*, **35**, 495–509.

Breeden, D. T. (1979). An intertemporal asset pricing model with stochastic consumption and investment opportunities, *Journal of Financial Economics*, **7**, 265–296.

Bruno, M. (1976). The two sector open economy and the real exchange rate, *American Economic Review*, **66**, 566–577.

Campbell, J. Y. (1987). Does saving anticipate declining labor income? An alternative test of the permanent income hypothesis, *Econometrica*, **55**, 1249–1273.

Campbell, J. Y. and A. Deaton (1989). Why is consumption so smooth? *Review of Economic Studies*, **56**, 357–73.

Campbell, J. H. and N. G. Mankiw (1990). Permanent income, current income, and consumption, *Journal of Business and Economic Statistics*, **8**, 265–279.

Cochrane, J. H. (1994). Permanent and transitory components of GNP and stock prices, *Quarterly Journal of Economics*, **109**, 241–265.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika*, **64**, 247–254.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika*, **74**, 33–43.

Deaton, A. (1987). Life-cycle models of consumption: Is the evidence consistent with the theory? *Advances in Econometrics, Fifth World Congress*, **2**, Cambridge and New York, Cambridge University Press.

Deaton, A. (1992). *Understanding Consumption*, Oxford, Clarendon Press.

Diebold, F. X. and G. D. Rudebusch (1991). Is consumption too smooth? Long memory and the Deaton paradox, *Review of Economics and Statistics*, **73**, 1–9.

Elwood, S. K. (1998). Is the persistence of shocks to output asymmetric? *Journal of Monetary Economics*, **41**, 411–426.

Engel, C. and J. D. Hamilton (1990). Long swings in the dollar: Are they in the data and do markets know it? *American Economic Review*, **80**, 689–713.

Engle, R.F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, *Econometrica*, **50**, 987–1008.

Falk, B. and B. S. Lee (1998). The dynamic effects of permanent and transitory labor income on consumption, *Journal of Monetary Economics*, **41**, 371–387.

Flavin, M. A. (1981). The adjustment of consumption to changing expectations about future income, *Journal of Political Economy*, **89**, 974–1009.

Flavin, M. A. (1993). The excess smoothness of consumption: Identification and interpretation, *Review of Economic Studies*, **60**, 651–666.

Frenkel, J. A. and A. Razin (1996). *Fiscal Policies and Growth in the World Economy*, Third Edition, Cambridge: MIT Press.

Friedman, M. (1957). *A Theory of Consumption Function*, Princeton, Princeton University Press.

Glick, R. and K. Rogoff (1992). Global versus country-specific productivity shocks and the current account, *National Bureau of Economic Research*, Working Paper: 4140.

Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process, *Journal of Financial Economics*, **42**, 27–62.

Hall, R. E. and F. S. Mishkin (1982). The sensitivity of consumption to transitory income: Estimates from panel data on households, *Econometrica*, **50**, 461–480.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica*, **57**, 357–384.

Hamilton, J. D. (1994). *Time Series Analysis*, Princeton, NJ: Princeton University Press.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica*, **64**, 413–430.

Hansen, L. P., and T. J. Sargent (1980). Formulating and estimating dynamic linear rational expectations models, *Journal of Economic Dynamics and Control*, **2**, 7–46.

Hansen, L. P., and T. J. Sargent (1981). A note on Wiener-Kolmogorov prediction formulas for rational expectations models, *Economics Letters*, **8**, 255–260.

Heckman, J. J. (1974). Shadow prices, market wages and labor supply, *Econometrica*, **42**, 679–694.

Kuan, C. M., Y. L. Huang, and R. S. Tsay (2003). A component-driven model for regime switching and its empirical evidence, *Journal of Business and Economic Statistics*, forthcoming.

Modigliani, F. and R. Brumberg (1954). Utility Analysis and the Consumption Function. An Introduction of Cross-Section Data, in Kenneth K. Kurihara, ed., *Post Keynesian Economics*, 388–436, New Brunswick, N.J.: Rutgers University Press.

Lee, B. S. (1995). The response of stock price to permanent and temporary shocks to dividends, *Journal of Financial and Quantitative Analysis*, **30**, 1–22.

Lee, B. S. (1996). Time-series implications of aggregate dividend behavior, *Review of Financial Studies*, **9**, 589–618.

Lintner, J. (1956). Distribution of incomes of corporations among dividends, retained earnings and taxes, *American Economic Review Proceedings*, **66**, 97–113.

Ljung, G. M. and G. E. P. Box (1978). On a measure of lack of fit in time series models, *Biometrika*, **65**, 297–303.

Lucas, R. E., Jr. (1967). Optimal investment policy and the flexible accelerator, *International Economic Review*, **8**, 78–85.

Lucas, R. E., Jr. (1978). Asset pricing in an exchange economy, *Econometrica*, **46**, 1426–1445.

Lucas, R. E., Jr. (1981). *Studies in Business Cycle Theory*, Cambridge: MIT Press.

Lucas, R. E., Jr. and L. A. Rapping (1969). Real wages, employment and inflation, *Journal of Political Economy*, **77**, 721–754.

Quah, D. (1990). Permanent and transitory movements in labor income: An explanation for 'excess smoothness' in consumption, *Journal of Political Economy*, **98**, 449–475.

Razin, A. (1995). The dynamic optimizing approach to the current account: Theory and evidence in Kenen P. (Ed.), *Understanding Interdependence: The Macroeconomics of the Open Economy.* Princeton University Press. Princeton, NJ.

Sachs, J. D. (1981). The current account and macroeconomic adjustment in the 1970s, *Brookings Papers on Economic Activity*, **1**, 201–268.

Sahasaku, C. (1986). The U.S. evidence on optimal taxation over time, *Journal of Monetary Economics*, **18**, 251–275.

Sargent, T. (1978). Estimation of dynamic labor demand schedules under rational expectations, *Journal of Political Economy*, **86**, 1009–1044.

West, K. D. (1988). The sensitivity of consumption to news about income, *Journal of Monetary Economics*, **21**, 17–34.