FACULTEIT ECONOMIE
EN BEDRIJFSKUNDE

TWEEKERKENSTRAAT 2
B-9000 GENT
Tel. : 32 - (0)9 – 264.34.61
Fax. : 32 - (0)9 – 264.35.92

WORKING PAPER

On the stability of the excess sensitivity of aggregate consumption growth in the US

Gerdie Everaert, Lorenzo Pozzi and Ruben Schoonackers

January 2016
2016/917

D/2016/7012/01

On the stability of the excess sensitivity of aggregate consumption growth in the US

Gerdie Everaert[1], Lorenzo Pozzi*[2], and Ruben Schoonackers[3]

[1] *Ghent University & SHERPPA*

[2] *Erasmus University Rotterdam & Tinbergen Institute*

[3] *National Bank of Belgium & Ghent University*

January 11, 2016

Abstract

This paper investigates the degree of time variation in the excess sensitivity of aggregate consumption growth to anticipated aggregate disposable income growth using quarterly US data over the period 1953-2014. Our empirical framework contains the possibility of stickiness in aggregate consumption growth and takes into account measurement error and time aggregation. Our empirical specification is cast into a Bayesian state space model and estimated using Markov Chain Monte Carlo (MCMC) methods. We use a Bayesian model selection approach to deal with the non-regular test for the null hypothesis of no time variation in the excess sensitivity parameter. Anticipated disposable income growth is calculated by incorporating an instrumental variables estimation approach into our MCMC algorithm. Our results suggest that the excess sensitivity parameter in the US is stable at around 0.24 over the entire sample period.

JEL Classification: E21, C11, C22, C26

Keywords: Excess sensitivity, time-variation, consumption, income, MCMC, Bayesian model selection

*Corresponding author at: Department of Economics, P.O. Box 1738, 3000 DR Rotterdam, the Netherlands, Tel:+31 (10) 4081256, Email: pozzi@ese.eur.nl, Website: http://people.few.eur.nl/pozzi.

1 Introduction

Traditional permanent income and life cycle models of consumption predict that (log) real aggregate private consumption follows a random walk (see Hall, 1978). Empirical studies however have revealed that aggregate consumption growth is excessively sensitive to anticipated disposable income growth (see e.g., Campbell and Mankiw, 1989, 1990, 1991). The most common interpretations given to this observation are the occurrence of liquidity constraints (see e.g., Flavin, 1985; Deaton, 1991; Ludvigson, 1999) and the prevalence of precautionary and buffer stock savings motives (see e.g., Carroll, 1992; Ludvigson and Michaelides, 2001) which increase the weight given by consumers to current income in their consumption decisions.

In a number of empirical papers, the assumption that the excess sensitivity (ES) parameter is constant has been relaxed in favor of time-varying specifications (see e.g., Campbell and Mankiw, 1991; McKiernan, 1996; Bacchetta and Gerlach, 1997; Girardin et al., 2000; Peersman and Pozzi, 2007). Some of these studies have reported that ES has become less important during the last decades in the US (see e.g., Bacchetta and Gerlach, 1997) and in other developed economies (see e.g., Girardin et al., 2000; Blundell-Wignall et al., 1991). This is attributed to financial liberalization and the development of financial markets. These structural developments are thought to have improved the possibilities of consumers to smooth consumption over time and across states of the world, i.e., by curbing the importance of credit constraints and precautionary saving motives in consumer decisions these developments have, over time, reduced the ES parameter.

Some recent papers, which belong to a different branch of the ES literature that implicitly assumes that the ES parameter is constant, argue that the measured degree of ES of aggregate consumption growth, while still statistically significant, becomes of lower magnitude once other forms of aggregate consumption predictability are taken into account (see e.g., Basu and Kimball, 2002; Sommer, 2007; Kiley, 2010; Carroll et al., 2011). Sommer (2007) and Carroll et al. (2011), in particular, show that the magnitude of the degree of ES measured in quarterly US data is considerably lower in a model that contains a mechanism - i.e., habit formation, rational inattention or imperfect information - that generates dependency of aggregate consumption growth to its own past (i.e., 'stickiness'). Sommer (2007) further argues that it is necessary to adequately deal with both measurement error and time aggregation to obtain valid estimates of the ES parameter.

The contribution of this paper to the literature is to link both these strands of research by investigating the potential *time variation* of the ES of aggregate consumption growth to anticipated disposable income growth using an empirical framework that contains the possibility of stickiness in aggregate consumption growth and that provides an adequate treatment of measurement error and time aggregation. As Gali et al. (2007) show that different sources and degrees of aggregate consumption growth predictability have different macroeconomic implications, it is important to correctly measure the potentially time-varying degree of ES using an appropriate empirical framework. Our empirical framework is applied to US data over the period $1953-2014$. The paper further contributes to the literature by suggesting an appropriate methodological approach that allows to test whether the time variation in the ES parameter is empirically relevant and that adequately deals with all the complications that arise when estimating time-varying ES in our elaborate empirical set-up. More specifically, our methodological approach is centered around three issues.

First, a key question is whether time variation in the ES parameter, which is modelled as a standard unobserved random walk process, is statistically relevant. This is a non-regular testing problem as the null hypothesis that the variance of the innovations to the time-varying ES parameter is zero lies on the boundary of the parameter space. No previous study that investigates the possible time variation in the ES parameter has taken this complication into account. We use the Bayesian model selection approach for state space models recently suggested by Frühwirth-Schnatter and Wagner (2010) to test for time variation in the ES parameter. Their approach applied to our time-varying parameter case implies splitting the time-varying parameter in a constant part and in a time-varying part and introducing a stochastic binary model indicator which is one if the time-varying part is to be included in the model and zero otherwise. Using Markov Chain Monte Carlo (MCMC) methods (i.e., Gibbs sampling), these stochastic binary indicators are then sampled jointly with the other model parameters. Moreover, for the variance of the innovation to the ES parameter, we do not use the standard inverse gamma (IG) prior that is usually employed for variance parameters in a Bayesian setting. Rather, we use a Gaussian prior centered at zero for *the standard deviation* of the innovation to the ES parameter. The reason for this is that when using an IG prior distribution for the variance parameters, the choice of the shape and the scale hyperparameters that define this distribution has a strong influence on the posterior distribution when the true value of the variance is close to zero. More specifically, as the IG distribution does not have probability mass at zero, using it as a prior distribution tends to push the posterior density away from zero. This is of particular importance when estimating the variance of the innovations to the time-varying ES parameter as the purpose of the paper is to decide whether time variation is relevant or not. An interesting implication of estimating the standard deviation instead of the variance of the the innovations to the time-varying parameters is that the sign of the standard deviation is not identified. This offers an extra piece of information as it implies that the posterior distribution becomes bimodal when there is time variation, while it is unimodal around zero when there is no time variation.

Second, as anticipated aggregate income growth is not observed, the ES parameter is estimated from the relation between consumption growth and ex-post observed aggregate disposable income growth using an instrumental variables (IV) method. Time-varying ES parameter models with endogenous regressors have been estimated by, among others, Bacchetta and Gerlach (1997) and Peersman and Pozzi (2007) using *approximate* methods. In a recent paper, Kim and Kim (2011) show that the control function approach to IV estimation can be used to construct an *exact* state space representation which can then be estimated by maximum likelihood (ML). Building on their paper, we incorporate a control function type of approach in our MCMC algorithm to deal with endogeneity. The advantage of our Gibbs sampling approach compared to ML estimation is that it is computationally easier to implement and, as such, does not suffer from the numerical optimization problems inherent to ML estimation (see Kim and Kim (2011)). Moreover, as our Bayesian IV approach relies on sampling the posterior distribution rather than on using asymptotic approximations, it allows for exact inference even when instruments are weak.

Third, the potential presence of stickiness in aggregate consumption growth, on the one hand, and time aggregation and measurement error in the log level of consumption, on the other hand, implies that aggregate consumption growth follows an autoregressive (AR) process with moving average (MA) errors, i.e., an AR(1) process with MA(3) errors. To obtain valid estimates for the ES parameter, the AR(1) term in the consumption growth equation and the MA terms in the error term of the consumption growth equation must be taken into account explicitly. To deal with the MA components in the error term, we follow Chib and Greenberg (1994) who present exact methods to analyze Bayesian regression models with MA errors using MCMC sampling.

Our estimation results suggest that the degree of excess sensitivity of aggregate consumption growth to anticipated disposable income growth for the US is stable and lies around 0.24 over the entire sample period (1954-2014). This estimated magnitude of the excess sensitivity parameter is in accordance with recent findings of Sommer (2007) and Carroll et al. (2011) who consider an empirical framework with a constant ES parameter but with the possibility of stickiness in aggregate consumption growth along with time aggregation and measurement error. The lack of time variation in the ES parameter however stands in contrast to the findings reported by some of the previous studies that investigate time-varying ES for the US (e.g., the study by Bacchetta and Gerlach (1997) who argue that the degree of ES has dropped gradually over time). The evidence in favor of time variation that is reported in the literature has been obtained from the estimation of more restricted empirical models however. These typically do not allow for stickiness in aggregate consumption growth, nor do they allow for time aggregation and measurement error. Upon estimating more restricted empirical models that are in line with the time-varying ES frameworks employed in previous studies, we do obtain evidence that is more supportive - albeit hardly conclusive - of time variation in ES. Our results therefore imply that the finding of time variation in the ES parameter may be due to specification errors and may not be the result of genuine structural economic developments such as financial liberalization. Our findings further confirm that there is notable stickiness in aggregate consumption growth as we find a coefficient on lagged consumption growth that lies around 0.55, a result which is in accordance with the findings reported recently by Sommer (2007), Kiley (2010) and Carroll et al. (2011).

The remainder of this paper is organized as follows. In section 2 we present the benchmark theoretical model for aggregate consumption growth to which we add anticipated disposable income growth to allow for time-varying excess sensitivity. Section 3 outlines our empirical specification and estimation methodology. Section 4 presents the estimation results for the US over the period 1953-2014. Section 5 concludes.

2 Theoretical framework

In this section we first present a benchmark model for aggregate consumption growth with stickiness modeled through habit formation in consumer preferences[1] and MA(3) errors to capture the effects of time aggregation and measurement error. This benchmark model is a generalization of the model presented by Sommer (2007) as it includes a time-varying intercept in aggregate consumption growth that allows to capture and control for unspecified and/or hard-to-estimate components of aggregate consumption growth. Following, among others, Bacchetta and Gerlach (1997) and Carroll et al. (2011) we then add anticipated income growth to the consumption growth equation to allow for time-varying ES of consumption to income.

2.1 A benchmark theoretical model with habit formation

Suppose a representative permanent income consumer maximizes the following stream of discounted utilities

$$\max E_t \sum_{j=0}^{T} \rho^j U\left(\overline{C}_{t+j}; X_{t+j}\right), \tag{1}$$

subject to a budget constraint, where E_t denotes the consumer's expectation conditional on period t information, ρ is the discount factor, $\overline{C}_t$ is the level of period t 'effective' consumption and X_t is a variable or a combination of variables that shifts marginal utility at time t.[2] 'Effective' consumption is

[1] Alternative mechanisms by which stickiness can be incorporated into aggregate consumption growth are rational inattention (see e.g., Reis, 2006; Carroll et al., 2011) and imperfect information (see e.g., Pischke, 1995).

[2] Examples are hours worked (see e.g., Kiley, 2010) and/or government consumption (see e.g., Evans and Karras, 1998).

assumed to be equal to

$$\overline{C}_t = C_t^* - \gamma C_{t-1}^*, \tag{2}$$

where C_t^* is the representative agent's consumption level in period t such that utility depends on the level of consumption C_t^* relative to last period's consumption level C_{t-1}^* with the parameter γ (where $0 \leq \gamma \leq 1$) capturing the strength of habits. When $\gamma = 0$ habits are irrelevant and the consumer derives utility only from the level of consumption. When $\gamma = 1$ habits are most important and the consumer derives utility only from the change in consumption. When $0 < \gamma < 1$ the consumer derives utility both from the level of consumption and from the change in consumption. Hayashi (1985) and Dynan (2000) show that, provided the real interest rate is constant and T is large, the first-order condition under time-nonseparable preferences can be written as

$$E_{t-1}\left[R\rho \frac{U'(\overline{C}_t; X_t)}{U'(\overline{C}_{t-1}; X_{t-1})}\right] = 1, \tag{3}$$

where R is the real interest factor (which equals 1 plus the real interest rate) and where $U'(\overline{C}_t; X_t) = \frac{\partial U(\overline{C}_t; X_t)}{\partial \overline{C}_t}$.

Assuming that the utility function is of the CRRA type, i.e., $U(\overline{C}_t; X_t) = \frac{\overline{C}_t^{1-\psi}}{1-\psi} X_t$ with $\psi > 0$, so that $U'(\overline{C}_t; X_t) = \overline{C}_t^{-\psi} X_t$ and using this into equation (3) gives

$$E_{t-1}\left[R\rho \left(\frac{\overline{C}_t}{\overline{C}_{t-1}}\right)^{-\psi} \frac{X_t}{X_{t-1}}\right] = E_{t-1}\left[Z_t\right] = 1, \tag{4}$$

where $Z_t \equiv R\rho \left(\frac{\overline{C}_t}{\overline{C}_{t-1}}\right)^{-\psi} \frac{X_t}{X_{t-1}}$. Assuming that $\Delta \ln \overline{C}_t$ and $\Delta \ln X_t$ are jointly conditionally normally distributed implies that $\ln Z_t = \ln(R\rho) - \psi \Delta \ln \overline{C}_t + \Delta \ln X_t$ is conditionally Gaussian as well. From the lognormal property we can therefore write

$$E_{t-1}\left[Z_t\right] = \exp\left[E_{t-1}(\ln Z_t) + \frac{1}{2} V_{t-1}(\ln Z_t)\right]. \tag{5}$$

We then substitute equation (5) into equation (4) and take logs of the resulting equality to obtain (after some rearrangements of terms)

$$E_{t-1}\left(\Delta \ln \overline{C}_t\right) = \frac{1}{2\psi} \sigma^2_{\ln Z,t} + \frac{1}{\psi} \ln(R\rho) + \frac{1}{\psi} \mu_{\Delta \ln X,t}, \tag{6}$$

where $\sigma^2_{\ln Z,t} \equiv V_{t-1}[\ln Z_t] = V_{t-1}[\Delta \ln X_t] - 2\psi cov_{t-1}[\Delta \ln X_t, \Delta \ln \overline{C}_t] + \psi^2 V_{t-1}[\Delta \ln \overline{C}_t]$ and where

$\mu_{\Delta \ln X,t} = E_{t-1}(\Delta \ln X_t)$. This can also be written as

$$\Delta \ln \overline{C}_t = \frac{1}{2\psi}\sigma^2_{\ln Z,t} + \frac{1}{\psi}\ln(R\rho) + \frac{1}{\psi}\mu_{\Delta \ln X,t} + \epsilon_t, \tag{7}$$

where $\epsilon_t = \left[\Delta \ln \overline{C}_t - E_{t-1}\left(\Delta \ln \overline{C}_t\right)\right]$. Note that the innovation ϵ_t implicitly reflects the revision in permanent income of the optimizing permanent income consumer (see e.g., Campbell and Mankiw, 1990).

After collecting the first three terms of equation (7) into a time-varying variable β_{0t} and using the approximation $\Delta \ln \overline{C}_t = \Delta \ln(C^*_t - \gamma C^*_{t-1}) \approx \Delta \ln C^*_t - \gamma \Delta \ln C^*_{t-1}$ as suggested by Muellbauer (1988) and Dynan (2000), we obtain

$$\Delta \ln C^*_t = \beta_{0t} + \gamma \Delta \ln C^*_{t-1} + \epsilon_t, \tag{8}$$

where $\beta_{0t} = \frac{1}{2\psi}\sigma^2_{\ln Z,t} + \frac{1}{\psi}\ln(R\rho) + \frac{1}{\psi}\mu_{\Delta \ln X,t}$. As such, β_{0t} is a catch-all term that allows to capture and control for unspecified (i.e., the conditional mean and variance of $\Delta \ln X_t$) and/or hard-to-estimate (i.e., the conditional variance $V_{t-1}[\Delta \ln \overline{C}_t]$ in $\sigma^2_{\ln Z,t}$) components of aggregate consumption growth.[3,4]

2.2 Time aggregation and measurement error

Assuming that consumption decisions are made more frequently than the intervals at which consumption is measured causes time aggregation. Sommer (2007) shows that time aggregation in combination with the presence of habits induces an MA(2) structure in the error term ϵ_t of true aggregate consumption growth $\Delta \ln C^*_t$, where 'true' refers to consumption in the absence of measurement error and other transitory components. This implies that equation (8) should be written as

$$\Delta \ln C^*_t = \beta_{0t} + \gamma \Delta \ln C^*_{t-1} + \theta^{\xi}(L)\,\xi_t, \tag{9}$$

with ξ_t an *i.i.d.* error term and $\theta^{\xi}(L) = 1 + \theta^{\xi}_1 L + \theta^{\xi}_2 L^2$ an MA(2) lag polynomial with parameters being complicated functions of γ.[5]

Sommer (2007) further notes that aggregate consumption data measured at the quarterly level are often plagued by measurement error and other sources of transitory consumption fluctuations. He argues that measurement error is best modeled as an MA(1) structure in the log-level of consumption. This implies that measured aggregate consumption growth $\Delta \ln C_t$ should be modeled as the sum of of true

[3]Note that in Sommer (2007) the intercept in consumption growth is assumed to be constant, i.e., $\beta_{0t} = \beta_0$ (for all t) as his model includes no preference shifters X_t while he implicitly assumes a constant conditional variance of consumption growth.

[4]Note that while the model is derived under a constant interest factor R, the presence of β_{0t} in the model implicitly also allows to control for a time-varying interest rate in the estimation.

[5]Note that the proof in Sommer (2007) is based on a model with a constant mean in aggregate consumption growth whereas we have the time-varying variable β_{0t} in the model. We can however rewrite equation (8) as $(\Delta \ln C^*_t - \mu_t) = \gamma(\Delta \ln C^*_{t-1} - \mu_{t-1}) + \epsilon_t$ with $\mu_t = \beta_{0t} + \gamma\mu_{t-1}$ so that his proof can equally be applied to our model.

aggregate consumption growth $\Delta \ln C_t^*$ given by equation (9) and an MA(2) error term

$$\Delta \ln C_t = \Delta \ln C_t^* + \theta^{\zeta}(L)\,\varsigma_t, \tag{10}$$

where ς_t an *i.i.d.* error term and $\theta^{\zeta}(L) = 1 + \theta_1^{\zeta} L + \theta_2^{\zeta} L^2$ an MA(2) lag polynomial. This specification assumes "general" measurement error. Note that this specification encompasses the simpler case where measurement error is assumed to be a white noise error term in the log-level of consumption. In that case, there is only an MA(1) error term in equation (10) where $\theta_1^{\zeta} = -1$. This is the case of "classical" measurement error.

Combining equations (9) and (10) to obtain an expression containing only measured consumption growth $\Delta \ln C_t$ gives

$$\begin{aligned} \Delta \ln C_t &= \beta_{0t} + \gamma \Delta \ln C_{t-1} + \theta^{\zeta}(L)\left(\varsigma_t - \gamma \varsigma_{t-1}\right) + \theta^{\xi}(L)\,\xi_t, \\ &= \beta_{0t} + \gamma \Delta \ln C_{t-1} + \theta(L)\,\varepsilon_t, \end{aligned} \tag{11}$$

with ε_t an *i.i.d.* error term and $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \theta_3 L^3$ an MA(3) lag polynomial.[6]

2.3 Time-varying excess sensitivity

Empirical studies have demonstrated that aggregate consumption growth is excessively sensitive to anticipated disposable income growth (see e.g., Campbell and Mankiw, 1989, 1990, 1991). We follow the standard approach to test for this type of ES by adding *anticipated* income growth to the consumption growth model. In line with, among others, Bacchetta and Gerlach (1997) we allow the ES parameter to vary over time. More specifically, we extend our benchmark in equation (11) to

$$\Delta \ln C_t = \beta_{0t} + \beta_{1t} E_{t-1}(\Delta \ln Y_t) + \gamma \Delta \ln C_{t-1} + \theta(L)\,\varepsilon_t, \tag{12}$$

where $\Delta \ln Y_t$ is aggregate (total) disposable income growth and β_{1t} is the time-varying ES parameter.[7] Note that as the innovation ε_t implicitly reflects the revision in permanent income of the optimizing permanent income consumer of sections 2.1 and 2.2, it can be correlated with the variable $\Delta \ln Y_t$ (see e.g., Campbell and Mankiw, 1990). We explicitly deal with this issue when estimating equation (12) as discussed below.

[6]Note that $\theta(L)\,\varepsilon_t$ is the sum of the two independent MA processes $\theta^{\zeta}(L)\left(\varsigma_t - \gamma\varsigma_{t-1}\right)$ and $\theta^{\xi}(L)\,\xi_t$, the first being of order order 3 and the second of order 2. Following Hamilton (1994), we can write the sum of two independent MA processes as an MA process of which the order equals that of the highest order process in the sum.

[7]This approach differs from the method followed by among others Campbell and Mankiw (1990) and Kiley (2010) who test for ES by writing down aggregate consumption growth as the sum of or (weighted) average between consumption growth of optimizing permanent income consumers and consumption growth of current income ('rule-of-thumb') consumers.

We use the model in equation (12) to test for time-varying ES of aggregate consumption growth with respect to income growth against the benchmark model in equation (11). This deviates from the past literature in a number of ways. First, ES is usually tested against a framework where aggregate consumption growth is either white noise (i.e., the standard random walk model) or an MA(1) process if time aggregation and classical measurement error are taken into account (see e.g., Bacchetta and Gerlach (1997)). The results of Sommer (2007) and Carroll et al. (2011) show however that allowing for stickiness (i.e., the dependence of aggregate consumption growth on its own lag) is important when testing for the ES of consumption to income. In our model this is incorporated by introducing a habit formation mechanism. Second, as noted by Sommer (2007), allowing for classical measurement error may not be sufficient such that a more general framework with MA(q) errors is called for. The relevant order q will be determined empirically. Third, the time-varying variable β_{0t} controls for all potentially omitted variables that may affect aggregate consumption growth (i.e., marginal utility shifters such as hours worked and government consumption, the conditional variance of consumption growth which reflects a potential precautionary savings motive, the interest rate that captures potential inter-temporal substitution effects).

3 Empirical methodology

In this section we outline our empirical specification and econometric methodology to estimate the model for aggregate consumption growth outlined in section 2.

3.1 Empirical specification

We use equation (12) to test for time-varying ES of aggregate consumption growth with respect to income growth against the benchmark model in equation (11). The empirical implementation of equation (12) requires a number of further assumptions. These are outlined below.

Time-varying parameters

The parameters β_{0t} and β_{1t} in equation (12) are allowed to change over time according to a random walk process

$$\beta_{i,t+1} = \beta_{it} + \eta_{it}, \qquad \eta_{it} \sim i.i.d.\mathcal{N}(0, \sigma^2_{\eta_i}), \tag{13}$$

with $i = 0, 1$. Random walk processes allow for a very flexible evolution of the parameters β_{it} over time.[8]

[8]As a robustness check, we have also estimated the time-varying intercept β_{0t} as an $AR(1)$ process. This alternative modeling strategy does not affect the conclusions reported in the paper.

Anticipated income growth and instrumental variables

Anticipated income growth $E_{t-1}(\Delta \ln Y_t)$ is not observed, but can be estimated by assuming that observed income growth $\Delta \ln Y_t$ is linearly related to a set of forecasting variables Z_t known to the consumer at time $t-1$, so that

$$\Delta \ln Y_t = Z_t \delta + \nu_t, \tag{14}$$

where Z_t is uncorrelated with ε_t in equation (12) and where ν_t an i.i.d. error term that is unpredictable at time t, i.e., $E_{t-1}\nu_t = 0$. Taking expectations E_{t-1} of equation (14) gives

$$E_{t-1}(\Delta \ln Y_t) = Z_t \delta + E_{t-1}\nu_t = Z_t \delta, \tag{15}$$

and substituting this in equation (12) yields

$$\Delta \ln C_t = \beta_{0t} + \beta_{1t} Z_t \delta + \gamma \Delta \ln C_{t-1} + \theta(L)\,\varepsilon_t. \tag{16}$$

Equation (16) is an instrumental variables (IV) type of regression model with instruments Z_t where δ is estimated using the first stage regression model (14). Because, as noted in section 2.3, the shocks to aggregate income growth and aggregate consumption growth are correlated, equations (14) and (16) are seemingly-unrelated regression equations with cross-equation parameter restrictions. The correlation structure in the error terms ν_t and ε_t can be expressed as

$$\Sigma_{\nu,\varepsilon} = \begin{bmatrix} \sigma_\nu^2 & \rho\sigma_\nu\sigma_\varepsilon \\ \rho\sigma_\nu\sigma_\varepsilon & \sigma_\varepsilon^2 \end{bmatrix}, \tag{17}$$

where ρ is the correlation between ν_t and ε_t. Using a Cholesky factorization of $\Sigma_{\nu,\varepsilon}$ we can write

$$\begin{bmatrix} \nu_t \\ \varepsilon_t \end{bmatrix} = \begin{bmatrix} \sigma_\nu & 0 \\ \rho\sigma_\varepsilon & \sigma_\varepsilon\sqrt{1-\rho^2} \end{bmatrix} \begin{bmatrix} \mu_{1t} \\ \mu_{2t} \end{bmatrix}, \tag{18}$$

where μ_{1t} and μ_{2t} are *i.i.d.* error terms with unit variance. Replacing ε_t in equation (16) by

$$\varepsilon_t = \frac{\rho\sigma_\varepsilon}{\sigma_\nu}\nu_t + \sigma_\varepsilon\sqrt{1-\rho^2}\mu_{2t}, \tag{19}$$

yields

$$\Delta \ln C_t = \beta_{0t} + \beta_{1t} Z_t \delta + \gamma \Delta \ln C_{t-1} + \rho\nu_t^* + \theta(L)\,\mu_t, \tag{20}$$

with $\nu_t^* = \sigma_\varepsilon \theta(L) \nu_t / \sigma_\nu$ and where $\mu_t = \sigma_\varepsilon \sqrt{1-\rho^2} \mu_{2t}$ is an *i.i.d.* error term that is not correlated with any other error term in the model. Note that, as a result, we have $\sigma_\varepsilon^2 = \sigma_\mu^2 / (1-\rho^2)$. Equation (20) is a control function type of IV regression model similar to the one outlined by Kim and Kim (2011) to deal with endogeneity in a time-varying parameter model.[9] However, instead of using their two-step or joint maximum likelihood (ML) procedure to estimate the non-linear model implied by equations (14) and (20), we use the Gibbs sampler as outlined in section 3.3 below. The advantage of our modeling and sampling approach compared to Kim and Kim (2011)'s two-step ML approach is that when estimating equation (20) we explicitly take into account that the error terms ν_t and ε_t may be correlated and that δ is estimated in a first step, so that $Z_t\widehat{\delta}$ is a generated regressor. The advantage of our modeling and sampling approach compared to Kim and Kim (2011)'s joint ML approach is that it is computationally easier to implement. As such, it does not suffer from the numerical optimization problems inherent to the joint ML estimation that are reported by Kim and Kim (2011). Moreover, as our Bayesian approach relies on sampling the posterior distribution rather than using asymptotic approximations, it allows for exact inference even when the instruments Z_t are weak.

3.2 Stochastic model specification search

A key question in the above model is whether the ES parameter β_{1t} is time-varying or constant. Although β_{1t} can be filtered using the Kalman filter and the variance of the innovations $\sigma_{\eta_1}^2$ can be estimated using ML, testing whether the time variation is relevant implies testing $\sigma_{\eta_1}^2 = 0$ against $\sigma_{\eta_1}^2 > 0$, which is a non-regular testing problem as the null hypothesis lies on the boundary of the parameter space. In a recent article, Frühwirth-Schnatter and Wagner (2010) show how to extend Bayesian model selection for standard regression models with observed variables to unobserved components in state space models. Their approach relies on a non-centered parameterization of the state space model in which (i) binary stochastic indicators for each of the model components are sampled together with the parameters and (ii) the standard inverse gamma (IG) prior for the variances of innovations to the components is replaced by a Gaussian prior centered at zero for the square root of these variances (i.e., for the standard deviations).

Non-centered parameterization

Frühwirth-Schnatter and Wagner (2010) argue that a first piece of information on the hypothesis whether the variance of innovations to a state variable is zero or not can be obtained by considering a non-centered parameterization. This implies rearranging the data generating process for the time-varying parameters

[9] An apparent difference is that our specification includes anticipated income growth, which is a predetermined regressor calculated using instrumental variables, while the specification of Kim and Kim (2011) includes an endogenous regressor. Besides the error term from the first step auxiliary regression, the control function equation in Kim and Kim (2011) therefore includes the endogenous regressor instead of the term $Z_t\delta$ that we include in our equation (20).

β_{it} in equation (13) to

$$\beta_{it} = \beta_{i0} + \sigma_{\eta_i}\beta^*_{it}, \tag{21}$$

$$\text{with} \qquad \beta^*_{i,t+1} = \beta^*_{it} + \eta^*_{it}, \qquad \beta^*_{i0} = 0, \qquad \eta^*_{it} \sim i.i.d.\mathcal{N}(0,1), \tag{22}$$

for $i = 0, 1$ and where β_{i0} is the initial value of β_{it} when this coefficient is time-varying ($\sigma_{\eta_i} > 0$) while being the constant value of β_{it} when there is no time variation ($\sigma_{\eta_i} = 0$). A crucial aspect of the non-centered parameterization is that it is not identified as the signs of σ_{η_i} and β^*_{it} can be changed by multiplying both with -1 without changing their product in equation (21). As a result of the non-identification, the likelihood function is symmetric around 0 along the σ_{η_i} dimension. When β_{it} is time-varying ($\sigma^2_{\eta_i} > 0$), the likelihood function is bimodal with modes $-\sigma_{\eta_i}$ and σ_{η_i}. For $\sigma^2_{\eta_i} = 0$, the likelihood function is unimodal around zero. As such, allowing for non-identification of σ_{η_i} provides useful information on whether $\sigma^2_{\eta_i} > 0$.

Stochastic model specification

A second advantage of the non-centered parameterization in equation (21) is that when $\sigma^2_{\eta_i} = 0$ the transformed component β^*_{it} (in contrast to β_{it}) degenerates to zero with the time-invariant parameter now represented by β_{i0}. As such, the question whether the ES parameter is time-varying or not can be expressed as a variable selection problem in equation (21). To this end, Frühwirth-Schnatter and Wagner (2010) introduce the stochastic model specification

$$\beta_{it} = \beta_{i0} + \iota_i\sigma_{\eta_i}\beta^*_{it}, \tag{23}$$

where ι_i is a binary indicator which is either 0 or 1. If $\iota_i = 0$, the component β^*_{it} drops from the model such that β_{i0} represents the constant intercept or slope parameter. If $\iota_i = 1$ then β^*_{it} is included in the model and σ_{η_i} is estimated from the data. In this case β_{i0} is the initial value of β_{it}.

Gaussian priors centered at zero for σ_{η_i}

Our Bayesian estimation approach requires choosing prior distributions for the model parameters. When using the standard IG prior distribution for the variance parameters, the choice of the shape and scale hyperparameters that define this distribution has a strong influence on the posterior distribution when the true value of the variance is close to zero. More specifically, as the IG distribution does not have probability mass at zero, using it as a prior distribution tends to push the posterior density away from zero. This is of particular importance when estimating the variances $\sigma^2_{\eta_i}$ of the innovations to the time-varying parameters β_{it} as for these components we want to decide whether they are relevant or not. As $\sigma^2_{\eta_i}$ is a regression coefficient in equation (23), a further important advantage of the non-centered parameterization is that it allows us to replace the standard IG prior on the variance parameter $\sigma^2_{\eta_i}$ by a Gaussian prior centered at zero on σ_{η_i}. Centering the prior distribution at zero makes sense as, for both $\sigma^2_{\eta_i} = 0$ and $\sigma^2_{\eta_i} > 0$, σ_{η_i} is symmetric around zero. Frühwirth-Schnatter and Wagner (2010) show that the posterior density of σ_{η_i} is much less sensitive to the hyperparameters of the Gaussian distribution and is not pushed away from zero when $\sigma^2_{\eta_i} = 0$. As such, we choose a Gaussian prior distribution centered at zero, i.e., $\mathcal{N}(0, V_0)$, for σ_{η_0} and σ_{η_1} where σ_{η_0} and σ_{η_1} are the standard deviations of the innovations to the time-varying parameters.

Other priors

For the variances of the error terms σ^2_μ and σ^2_ν, which are always included in the model, we choose the standard IG prior distribution $IG(c_0, C_0)$ where c_0 denotes the shape and C_0 denotes the scale of the distribution. For each of the model parameters β_{00}, β_{10}, γ, ρ, θ and δ, we assume a normal prior distribution $\mathcal{N}(b_0, V_0)$. Details on the chosen hyperparameters (b_0, V_0) for the prior $\mathcal{N}$ distributions and (c_0, C_0) for the prior IG distributions are presented in section 4.2 below. For the binary indicators ι_0 and ι_1 we choose a uniform prior distribution where each model component has a p prior probability of being included in the model, i.e., $p(\iota_0 = 1) = p(\iota_1 = 1) = p$.[10]

3.3 Gibbs sampler

Using equation (23), the model in equation (20) can be rewritten as

$$\Delta \ln C_t = (\beta_{00} + \iota_0 \sigma_{\eta_0} \beta^*_{0t}) + (\beta_{10} + \iota_1 \sigma_{\eta_1} \beta^*_{1t})\, Z_t \delta + \gamma \Delta \ln C_{t-1} + \rho \nu^*_t + \theta\,(L)\, \mu_t. \tag{24}$$

Taken together, equations (14) and (24) can be considered as the observation equations of a state space (SS) model, with the unobserved states β^*_{0t} and β^*_{1t} evolving according to the state equations in (22). In a standard linear Gaussian SS model, the Kalman filter can be used to filter the unobserved states from the data and to construct the likelihood function such that the unknown parameters can be estimated using maximum likelihood. However, the stochastic model specification search outlined in subsection 3.2 implies a non-regular estimation problem for which the standard approach via the Kalman filter and maximum likelihood is not feasible. Instead, we use the Gibbs sampler which is a Markov Chain Monte Carlo (MCMC) method to simulate draws from the intractable joint and marginal posterior distributions

[10]Note that when $p = 0.5$ (i.e., for our baseline case), each of the four models (i.e., the four combinations of the binary indicators) has the same prior probability equal to 0.25.

of the unknown parameters and the unobserved states using only tractable conditional distributions. Intuitively, this amounts to reducing the complex non-linear model into a sequence of blocks for subsets of parameters/states that are tractable conditional on the other blocks in the sequence.

For notational convenience, define the time-varying parameter vector $\beta_t^* = (\beta_{0t}^*, \beta_{1t}^*)$, the unknown parameter vectors $\phi_1 = (\delta, \sigma_\nu^2)$, $\phi_2 = (\beta_{00}, \beta_{10}, \sigma_{\eta_0}, \sigma_{\eta_1}, \gamma, \rho, \sigma_\mu^2)$ and $\phi = (\phi_1, \phi_2, \theta)$ and the model indicator $\mathcal{M} = (\iota_0, \iota_1)$.[11] Let $D_t = (\Delta \ln C_t, \Delta \ln Y_t, Z_t)$ be the data vector. Stacking observations over time, we denote $D = \{D_t\}_{t=1}^T$ and similarly for β^*. The posterior density of interest is then given by $f(\phi, \beta^*, \mathcal{M}|D)$. Building on Frühwirth-Schnatter and Wagner (2010) for the stochastic model specification part and on Chib and Greenberg (1994) for the moving average (MA) part, our MCMC scheme is as follows:

1. Sample the first step parameters $\phi_1 = (\delta, \sigma_\nu^2)$ from $f(\phi_1|D)$ using the regression model in equation (14) and calculate $Z_t\delta$ and ν_t. Then ν_t^* can be calculated using $\nu_t^* = \sigma_\varepsilon \theta(L) \nu_t / \sigma_\nu$.

2. Sample the MA coefficients θ from $f(\theta|\phi_1, \phi_2, \beta^*, \mathcal{M}, D)$ conditional on the parameters ϕ_2, the time-varying parameters β^*, the binary indicators in $\mathcal{M}$ and $Z_t\delta$ and ν_t^* calculated in the first block.

3. Sample the binary indicators $\mathcal{M}$ and the second step parameters ϕ_2 using the non-centered parameterization in equation (24) conditional on the MA coefficients θ, the time-varying parameters β^* and $Z_t\delta$ and ν_t^* calculated in the first block.

   (a) Sample the binary indicators $\mathcal{M}$ from $f(\mathcal{M}|\phi_1, \theta, \beta^*, D)$ marginalizing over the parameters ϕ_2 for which variable selection is carried out.

   (b) Sample the unrestricted parameters in ϕ_2 from $f(\phi_2|\phi_1, \theta, \beta^*, \mathcal{M}, D)$ while setting the restricted parameters σ_{η_i} (for which the corresponding component β_{it}^* is not included in the model $\mathcal{M}$) equal to 0.

4. Sample the unrestricted (i.e. for which $\iota_i = 1$) time varying parameters in β^* from $f(\beta^*|\phi, \mathcal{M}, D)$ again using the non-centered parameterization in equation (24) conditional on the second step parameters ϕ_2, the binary indicators $\mathcal{M}$ and $Z_t\delta$ and ν_t^* calculated in the first block. The restricted time varying parameters (for which $\iota_i = 0$) in β^* are sampled directly from their prior distribution using equation (22).[12]

[11] Note that $\theta = (\theta_1, \theta_2, ..., \theta_q)$.

[12] Even when $\iota_i = 0$ a sample for β_{it}^* is required as this will be used to calculate the marginal likelihood of a model with a time-varying β_{it}^* in block 3(a).

5. Perform a random sign switch for σ_{η_i} and $\{\beta^*_{it}\}_{t=1}^T$, i.e., σ_{η_i} and $\{\beta^*_{it}\}_{t=1}^T$ are left unchanged with probability 0.5 while with the same probability they are replaced by $-\sigma_{\eta_i}$ and $\{-\beta^*_{it}\}_{t=1}^T$.

Given an arbitrary set of starting values, sampling from these blocks is iterated J times and, after a sufficiently large number of burn-in draws B, the sequence of draws $(B+1, ..., J)$ approximates a sample from the virtual posterior distribution $f(\phi, \beta^*, \mathcal{M}|D)$. Details on the exact implementation of each of the blocks can be found in Appendix A. The results reported below are based on 10000 Gibbs sampler iterations, with the first 5000 draws discarded as a burn-in sequence.

4 Empirical results

4.1 Data

To estimate the empirical model, quarterly US data are available for all variables used over the period $1953Q1 - 2014Q4$. The use of lagged instruments reduces the sample size with 2 observations so that the effective sample period is $1953Q3 - 2014Q4$, i.e., this implies an effective sample size equal to 246 observations.[13] Where necessary, data are seasonally adjusted. For C_t, we use real per capita expenditures on nondurables and services (excluding clothing and footwear). For Y_t, real per capita personal disposable income is used. Both variables are put in real terms using the deflator of nondurables and services (excluding clothing and footwear) with base year $2009 = 100$. With respect to the estimation of anticipated income growth, external instruments are also used. In particular, we include as instruments lagged changes in the short run interest rate for which we take the three-month nominal T-bill rate, lagged changes in stock prices which we proxy using the S&P 500 index, lags of the inflation rate where the inflation rate is calculated as the log change in the CPI index, lags of the level of consumer confidence for which we take the index of consumer sentiment, and lags of the change in the unemployment rate.

Data for nominal expenditures on nondurables and services (excluding clothing and footwear), for nominal personal disposable income and for the corresponding deflator are taken from the National Product and Income Accounts (NIPA). Population data are taken from the OECD Quarterly National Accounts. For the three-month T-bill rate, data are taken from the Board of Governors. Data for the S&P 500 index comes from Sommer (2007) and is updated with data from Thomson Reuters Datastream. Finally, for the CPI index and the unemployment rate, data are taken from the Bureau of Labor Statistics while for the consumer sentiment index, the University of Michigan index is used.

[13] Note that for some variables instruments are formed using further lags (i.e., a third and fourth lag). As data for these variables are typically available well before $1953Q1$, these deeper lags do not reduce the effective sample size.

4.2 Prior choice

With respect to model selection, we mention that for the binary indicators ι_0 and ι_1 we choose a uniform prior distribution where, in the baseline case, each time-varying model component has a $p = 0.5$ prior probability of being included in the model.

Summary information on the prior distributions of the other unknown parameters is reported in Table 1. For the variances σ_μ^2 and σ_ν^2 of the error terms in the consumption and income growth equations, we use an inverse gamma prior distribution $IG(c_0, C_0)$ where the shape $c_0 = \nu_0 T$ and scale $C_0 = c_0\sigma_0^2$ parameters are calculated from the prior belief σ_0^2 and the prior strength ν_0, which is expressed as a fraction of the sample size T.[14] Our prior belief for σ_μ is 0.5, implying that 95% of the quarterly consumption growth shocks lie between -1% and 1%, while our prior belief for σ_ν of 0.75 implies that the 95% of the quarterly income growth shocks lie between -1.5% and +1.5%. The smaller value for σ_μ reflects the idea that income is more volatile than consumption. In both cases, the prior is fairly loose with strength set equal to 0.1.

For the remaining parameters, Gaussian prior distributions $\mathcal{N}(b_0, V_0)$ are used. First, consider the time-varying ES parameter, β_{1t}. For β_{10}, the prior is given by $\beta_{10} \sim \mathcal{N}(0.4, 0.2^2)$ which reflects our belief that if there is no time variation in β_{1t} (i.e., $\sigma_{\eta_1} = 0$) then the ES parameter ranges from roughly 0 to 0.8. This encompasses all values found in the literature. Campbell and Mankiw (1990) for example report values of 0.5 up to 0.7 for the U.S. Controlling for habits, Kiley (2010) and Sommer (2007) find lower values of about 0.3 and 0.15 respectively. For the standard deviation σ_{η_1} of the innovations to the time-varying part in β_{1t} a Gaussian prior centered at zero $\mathcal{N}(0, 0.2^2)$ is chosen. Note that the prior standard deviation $\sqrt{V_0} = 0.2$ implies a very loose prior as it allows that 95% of the standard deviations of the quarterly innovations to the ES parameter lie between -0.39 and 0.39.

For the time-varying intercept, β_{0t}, the prior distribution for the time-invariant part is fairly uninformative and centered at zero, $\beta_{00} \sim \mathcal{N}(0, 1)$. The prior belief $\sigma_{\eta_0} \sim \mathcal{N}(0, 0.2^2)$ about the degree of time-variation in β_{0t} is also centered at zero with the same prior standard deviation as the innovations to the ES parameter.

According to Carroll et al. (2011), the strength of habits in aggregate consumption growth for the U.S. varies between 0.5 and 0.7. These results are confirmed by, amongst others, Fuhrer (2000) and Sommer (2007) who both find a stickiness parameter around 0.7. Therefore our prior for γ is $\mathcal{N}(0.6, 0.15^2)$ such that the 95% prior interval ranges from roughly 0.3 to 0.9. For the MA parameters θ, a loose prior centered at zero is used.[15]

[14]Since this prior is conjugate, $\nu_0 T$ can be interpreted as the number of fictitious observations used to construct the prior belief σ_0^2.

[15]For the initial conditions λ of the MA process, a Gaussian prior with mean 0 and variance 1 is used (unreported in

Table 1: Prior distributions of model parameters

Inverse Gamma priors: $IG(c_0, C_0) = IG(\nu_0 T, \nu_0 T \sigma_0^2)$				Percentiles	
		σ_0	ν_0	2.5%	97.5%
error term consumption equation	σ_μ	0.50	0.10	0.42	0.62
error term income equation	σ_ν	0.75	0.10	0.63	0.94
Gaussian priors: $\mathcal{N}(b_0, V_0)$				Percentiles	
Non-centered components		b_0	$\sqrt{V_0}$	2.5%	97.5%
std. of time-varying intercept	σ_{η_0}	0.00	0.20	−0.39	0.39
std. of time-varying ES parameter	σ_{η_1}	0.00	0.20	−0.39	0.39
Model parameters					
constant value intercept	β_{00}	0.00	1.00	−1.96	1.96
constant value ES parameter	β_{10}	0.40	0.20	0.01	0.79
consumption habits parameter	γ	0.60	0.15	0.31	0.89
degree of correlation between ν_t and ε_t	ρ	0.00	0.40	−0.78	0.78
MA parameters	θ	0.00	0.50	−0.98	0.98
parameters first stage income equation	δ	0.00	0.50	−0.98	0.98

Notes: We set IG priors on the variance parameters σ^2 but in the top panel of this table we report details on the implied prior distribution for the standard deviations σ as these are easier to interpret. Likewise, in the bottom panel of the table we report $\sqrt{V_0}$ instead of V_0.

For the degree of correlation ρ between ν_t and ε_t, an uninformative prior is chosen, i.e., $\rho \sim \mathcal{N}(0, 0.4^2)$. A loose prior centered at zero is used for the parameters δ on the instrumental variables used to proxy $E_{t-1}(\Delta \ln Y_t)$.

4.3 Estimation results

We successively estimate five empirical models with increasing complexity. The fifth and last model coincides with the empirical specification presented in section 3. This approach facilitates the investigation of the impact of the features incorporated into the model on the obtained estimates for excess sensitivity and its variation over time. It also allows us to compare our findings to some of the findings reported in the literature. Hence, for each of the models, the importance of time variation in the ES parameter is discussed as is the robustness of the results under different instrument sets. We end this section with the presentation of the results for the parsimonious model that is selected by the stochastic model specification search.

Instrument sets

Since anticipated income growth $E_{t-1}(\Delta \ln Y_t)$ is not observed, a set of instrumental variables Z_t is used to estimate it. As Campbell and Mankiw (1990) and Kiley (2010), among others, show that the choice

Table 1). We refer to Appendix A for details on the estimation of the MA process.

of instruments can be critically important, the evaluation of the time variation in the ES parameter is reported using three different instrument sets. A first instrument set, Z^1, is based upon Campbell and Mankiw (1990) and includes a constant, lags 1-4 of disposable income growth and consumption growth, a lagged error correction term, i.e., log consumption minus log disposable income (see also Campbell, 1987)), lags 1-2 of changes in the short term interest rate and lags 1-2 of changes in stock prices. To construct the second instrument set, Z^2, we add the first and second lag of the inflation rate to the instruments contained in Z^1 (see Fuhrer, 2000; Kiley, 2010). Following Sommer (2007), the third instrument set Z^3 is constructed by adding lags 1-2 of the consumer sentiment index and of the change in the unemployment rate.[16,17]

In column 2 of Table 2 we report the average explanatory power of the different instrument sets in explaining $\Delta \ln Y_t$. For all instrument sets we find an average adjusted R^2 over all iterations of about 30%, which is very reasonable.

Model 1 (M1): no habits, no time-varying intercept, no MA components in the error term

We start by testing for time variation in the ES parameter using a basic model in which there is no time variation in the intercept ($\sigma_{\eta_0} = 0$), no dependency of aggregate consumption growth on its own past, and no MA structure in the error term. Based on equation (20), the empirical specification for aggregate consumption growth then becomes,

$$\Delta \ln C_t = \beta_{00} + \beta_{1t} Z_t \delta + \rho \nu_t^* + \mu_t,$$

where the data generating process for β_{1t} is represented by equations (22) and (23).

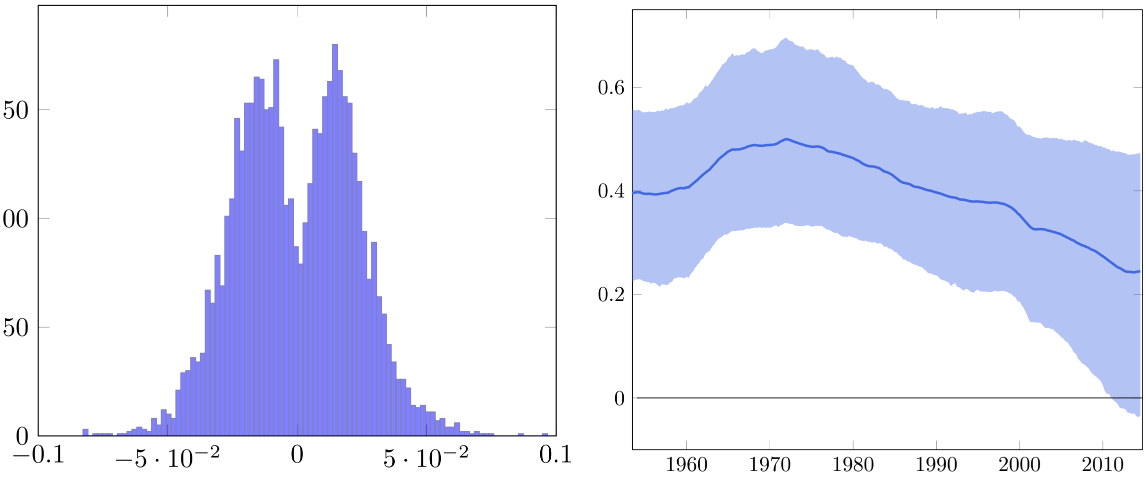
We first estimate this model with the binary indicator ι_1 set to 1 to generate a posterior distribution for the standard deviation (σ_{η_1}) of the innovations to the ES parameter. If this distribution is bimodal, with low or no probability mass at zero, this can be taken as a first indication of a time-varying ES. Figure 1 presents the resulting posterior distribution of σ_{η_1} as well as the mean of the posterior distribution of the time-varying ES parameter and its 90 % highest posterior density (HPD) interval. When looking at panel (a), we find clear-cut bimodality in the posterior distribution of the standard deviation of the innovations to the ES parameter, pointing to an important amount of time variation. Panel (b) further shows the resulting time variation in ES, which starts at a value close to 0.4 in 1953 and increases to

[16] Note that our results are very similar when we use lags 1 to 4 for the external instruments (i.e., for the instruments not obtained from disposable income and consumption) instead of lags 1 and 2.

[17] Concerning the choice of lags, note that, in general, the presence of autocorrelation of the MA form in the error term μ_t necessitates the use of instruments that are appropriately lagged, i.e., depending on the order of the MA component in μ_t. Our empirical approach, however, explicitly takes into account and controls for the MA terms so that we do not face this issue.

around 0.5 in the early 1970s after which it keeps on decreasing until it lies around 0.25 in 2014.

Figure 1: Stochastic model selection and time-varying parameters (binary indicators set to 1) in M1



(a) Posterior distribution σ_{η_1}

(b) Evolution time-varying ES parameter β_{1t}

Note: Figures are presented for the results using instrument set Z^3 but are similar when using instrument sets Z^1 and Z^2

As a more formal test for time variation, we next sample the stochastic binary indicator ι_1 together with the other parameters of the model. Table 2 reports the posterior probabilities that the binary indicators ι_i attached to the time-varying parameters β_{it} are equal to one for each of the five different models that we estimate and for the three instrument sets discussed above. The posterior probabilities for the binary indicators are calculated as the average selection frequencies over all iterations of the Gibbs sampler. In the baseline scenario, we assign a 0.5 prior probability to each of the binary indicators being one. Results for this baseline scenario are presented in the upper part of Table 2. As a sensitivity control, we re-estimate the different models with the prior inclusion probabilities set to 0.1 and 0.9 respectively. The resulting posterior probabilities are reported in the middle and lower part of Table 2.

For M1, the results in the baseline scenario ($p = 0.5$) show that when using instrument sets Z^2 and Z^3 the inclusion probability of a time-varying ES parameter is 0.27. For instrument set Z^1 it is somewhat lower. Only when increasing the prior inclusion probability to 0.9, there is some sign of time variation. All in all, the results indicate that, despite the bimodal posterior distribution of σ_{η_1} and a moderate reduction in the ES parameter over time, there is no real evidence in favor of time-varying excess sensitivity.

Table 2: Posterior inclusion probabilities for the binary indicators over different models and instrument sets

Prior	Instrument set		Posterior							
			M1	M2	M3		M4		M5	
	Z	R^2_{adj}	ι_1	ι_1	ι_0	ι_1	ι_0	ι_1	ι_0	ι_1
$p = 0.5$	Z^1	0.30	0.18	0.18	0.63	0.17	0.06	0.06	0.06	0.07
	Z^2	0.33	0.27	0.22	0.65	0.18	0.06	0.06	0.06	0.07
	Z^3	0.26	0.27	0.21	0.68	0.15	0.06	0.06	0.06	0.07
$p = 0.1$	Z^1	0.30	0.03	0.03	0.16	0.03	0.01	0.01	0.01	0.01
	Z^2	0.33	0.06	0.04	0.17	0.03	0.01	0.01	0.01	0.01
	Z^3	0.26	0.06	0.04	0.20	0.03	0.01	0.01	0.01	0.01
$p = 0.9$	Z^1	0.30	0.60	0.63	0.93	0.60	0.36	0.37	0.38	0.40
	Z^2	0.33	0.71	0.67	0.92	0.60	0.35	0.37	0.37	0.39
	Z^3	0.26	0.71	0.66	0.93	0.54	0.35	0.36	0.37	0.38
	Model specification									
	Habits		No	No	No		Yes		Yes	
	TV intercept		No	No	Yes		Yes		Yes	
	MA error terms		No	MA(1)	MA(1)		MA(2)		MA(3)	

Notes: The prior inclusion probability is given by $p = p(\iota_0 = 1) = p(\iota_1 = 1)$. The instrument set Z^1 includes lags 1-4 of disposable income growth and consumption growth, a lagged error correction term and lags 1-2 of the change in stock prices and the change in the short term interest rate. Instrument set Z^2 adds the first and second lag of the inflation rate. Instrument set Z^3 further includes lags 1-2 of the consumer sentiment index and of the change in the unemployment rate.

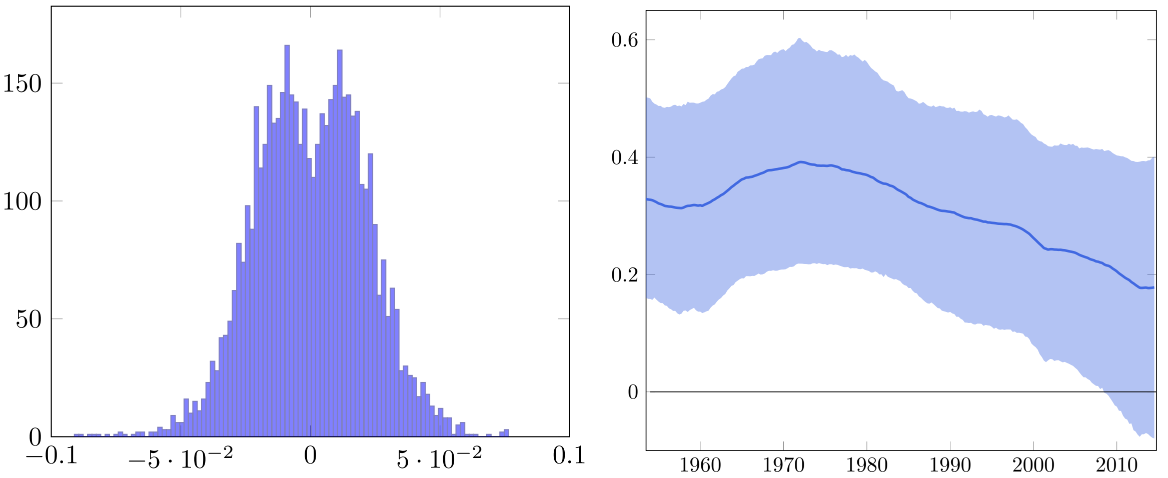
Model 2 (M2): no habits, no time-varying intercept, MA(1) error term

In the absence of habits, time aggregation and classical measurement error induce an MA(1) structure in the growth rate of consumption. This leads to the following empirical specification for aggregate consumption growth,

$$\Delta \ln C_t = \beta_{00} + \beta_{1t} Z_t \delta + \rho \nu_t^* + \theta(L)\mu_t,$$

with $\theta(L) = 1 + \theta_1 L$ an MA(1) lag polynomial. We again report the results in two steps. First, Figure 2 shows the posterior distribution for the standard deviation σ_{η_1} and plots the time-varying ES-parameter. Second, the individual posterior probability for the binary indicator ι_1 is reported in Table 2. Panel (a) of Figure 2 suggests that there is some bimodality in the posterior distribution for σ_{η_1} but compared to M1 it is less clear. Looking at the posterior inclusion probability for the time-varying part of the ES parameter in Table 2 shows that in the baseline scenario there is no significant time variation as for all instrument sets the probabilities vary around 0.2. For the two other scenarios (i.e., when $p(\iota_i = 1) = 0.1$ and when $p(\iota_i = 1) = 0.9$), posterior probabilities are also lower than the corresponding ones reported for M1.

Figure 2: Stochastic model selection and time-varying parameters (binary indicators set to 1) in M2



(a) Posterior distribution σ_{η_1} (b) Evolution time-varying ES parameter β_{1t}

Note: Figures are presented for the results using instrument set Z^3 but are similar when using instrument sets Z^1 and Z^2

Further, to underline the importance of controlling for the MA process in the error term, we report the posterior distribution of the different MA coefficients θ in Table 3. For M2, the 95% HPD interval of θ_1 varies between 0.04 and 0.33 which shows that controlling for MA terms is necessary.

A model similar to M2 is estimated by McKiernan (1996). In contrast to our analysis, their results indicate that the relationship between income and consumption is rather variable over the period 1959 – 1994 as a likelihood ratio test rejects the null hypothesis of a fixed parameter model against the alternative of a stochastic parameter model. However, similar to our results they do not find a notable decrease over time of the ES parameter.

Table 3: Posterior distribution for the MA parameters over different models

	M2	M3	M4		M5		
	θ_1	θ_1	θ_1	θ_2	θ_1	θ_2	θ_3
2.5%	0.04	0.03	-0.47	-0.12	-0.43	-0.10	-0.07
mean	0.19	0.17	-0.27	0.05	-0.22	0.07	0.08
97.5%	0.33	0.32	-0.05	0.23	0.05	0.25	0.23

Note: Results presented are based on instrument set Z^3 but are similar when using other instrument sets. Results are obtained with binary indicators set to 1.

Model 3 (M3): no habits, time-varying intercept, MA(1) error term

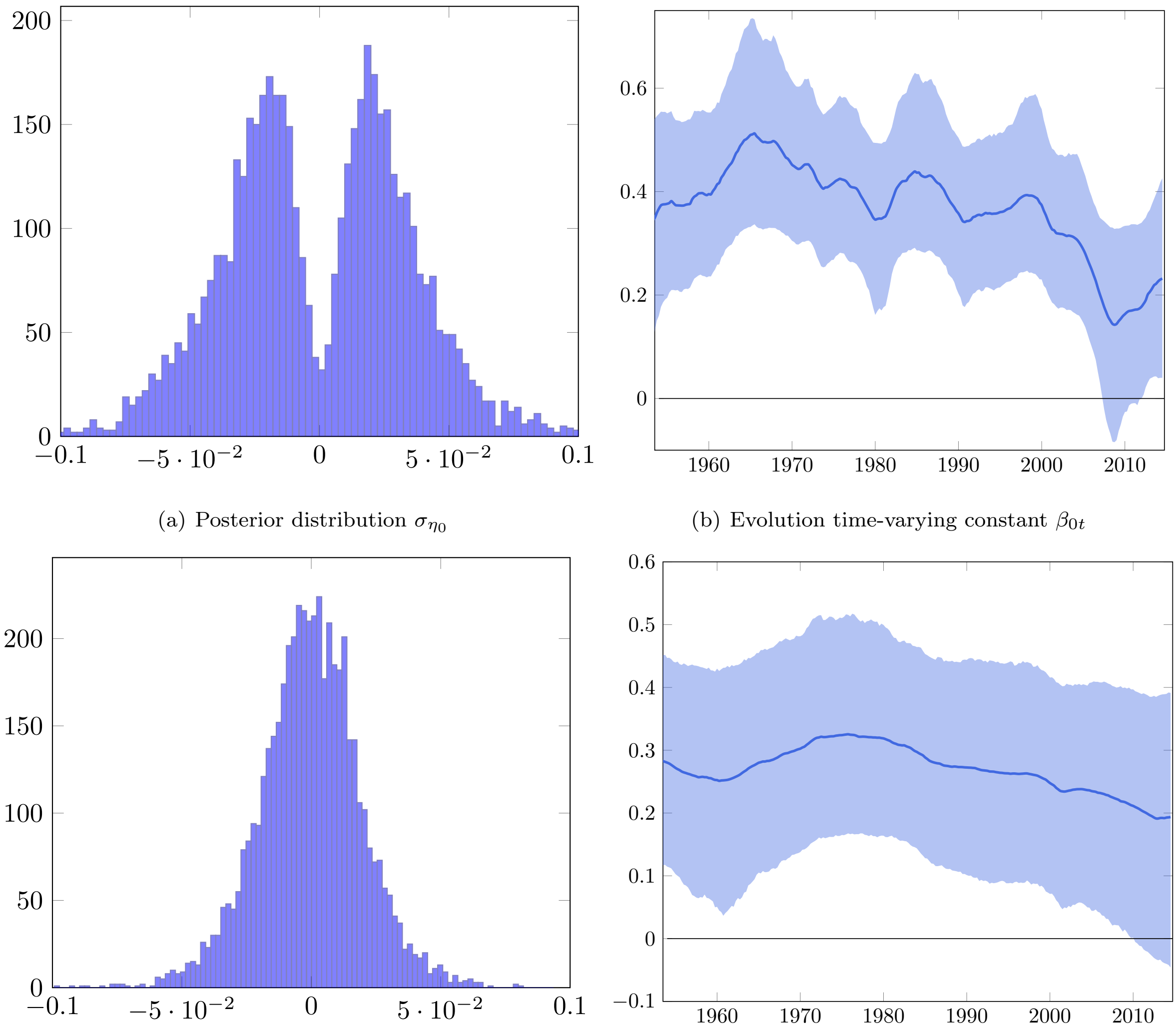
Models M1 and M2 are rather restrictive as they do not allow other variables, besides expected income growth, to have an impact on aggregate consumption growth. In M3 we allow for a time-varying constant β_{0t} that controls for potentially omitted variables that may affect aggregate consumption growth. The empirical specification for $\Delta \ln C_t$ then becomes,

$$\Delta \ln C_t = \beta_{0t} + \beta_{1t} Z_t \delta + \rho \nu_t^* + \theta(L)\mu_t,$$

with $\theta(L) = 1 + \theta_1 L$ an MA(1) lag polynomial and where equations (22) and (23) represent the processes for the time-varying variables β_{0t} and β_{1t}.

As we now allow for a time-varying intercept, Figure 3 and Table 2 also provide information on whether the time variation in the intercept is empirically relevant. When analyzing the posterior distribution of σ_{η_0}, we notice that Figure 3 panel (a) provides evidence of a bimodal distribution with low probability mass at zero, pointing to significant time-variation in the intercept. When analyzing the posterior distribution of σ_{η_1}, Figure 3 panel (c) suggests that there is no bimodality in the posterior distribution of σ_{η_1} and thus no time variation in ES. This is confirmed in Table 2 as in the baseline scenario the posterior probability of ι_1 equal to 1 is below 0.20 for all instrument sets. Even when increasing the prior inclusion probability up to 0.9, the posterior does not exceed 0.60. The model selection thus clearly rejects time variation in the ES parameter. For the intercept on the contrary, results are mixed. While Figure 3 panel (a) points to a time-varying intercept, results on the posterior inclusion probability give no clear evidence for time variation in the intercept. Related to the importance of taking into account the MA process in the error terms, Table 3 shows a similar posterior distribution for the MA coefficient in M3 as the one found for the MA parameter in M2.

Figure 3: Stochastic model selection and time-varying parameters (binary indicators set to 1) in M3



(a) Posterior distribution σ_{η_0}

(b) Evolution time-varying constant β_{0t}

(c) Posterior distribution σ_{η_1}

(d) Evolution time-varying ES parameter β_{1t}

Note: Figures are presented for the results using instrument set Z^3 but are similar when using instrument sets Z^1 and Z^2

Our results for the ES parameter as reported in Figure 3 panel (d) differ from the ES estimates reported by Bacchetta and Gerlach (1997) who estimate excess sensitivity in an empirical framework that is similar to our M3. They find that the ES of consumption to income falls gradually from about 0.75 in the early 1970s to about 0.4 in the early 1990s. Our findings in this framework (M3), on the other hand, suggest that ES has been relatively stable. We note that Bacchetta and Gerlach (1997) do not explicitly test for time variation in the ES parameter.

Models 4 and 5 (M4, M5): habits, time-varying intercept, MA(2)/MA(3) error term

Finally, in M4 and M5 we also allow for habits in aggregate consumption growth. The difference between both models is that M4 assumes 'classical' measurement error while M5 assumes general' measurement error. We refer to section 2.2 for details.[18] The empirical specification for $\Delta \ln C_t$ is given by,

$$\Delta \ln C_t = \beta_{0t} + \beta_{1t} Z_t \delta + \gamma \Delta \ln C_{t-1} + \rho \nu_t^* + \theta(L)\mu_t,$$

where for M4 we have $\theta(L) = 1 + \theta_1 L + \theta_2 L^2$ while for M5 we have $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \theta_3 L^3$ and where the data generating processes for β_{0t} and β_{1t} are shown by equations (22) and (23). Note that M5 coincides with the full empirical specification in equation (20). The results obtained for both models are almost identical. This is due to the fact that the posterior distribution of the MA(3) parameter θ_3 has considerable probability mass at zero, as reported in Table 3, making both models virtually indistinguishable. As such, we present only the graphs for M5 in Figure 4 as the ones for M4 are almost identical.

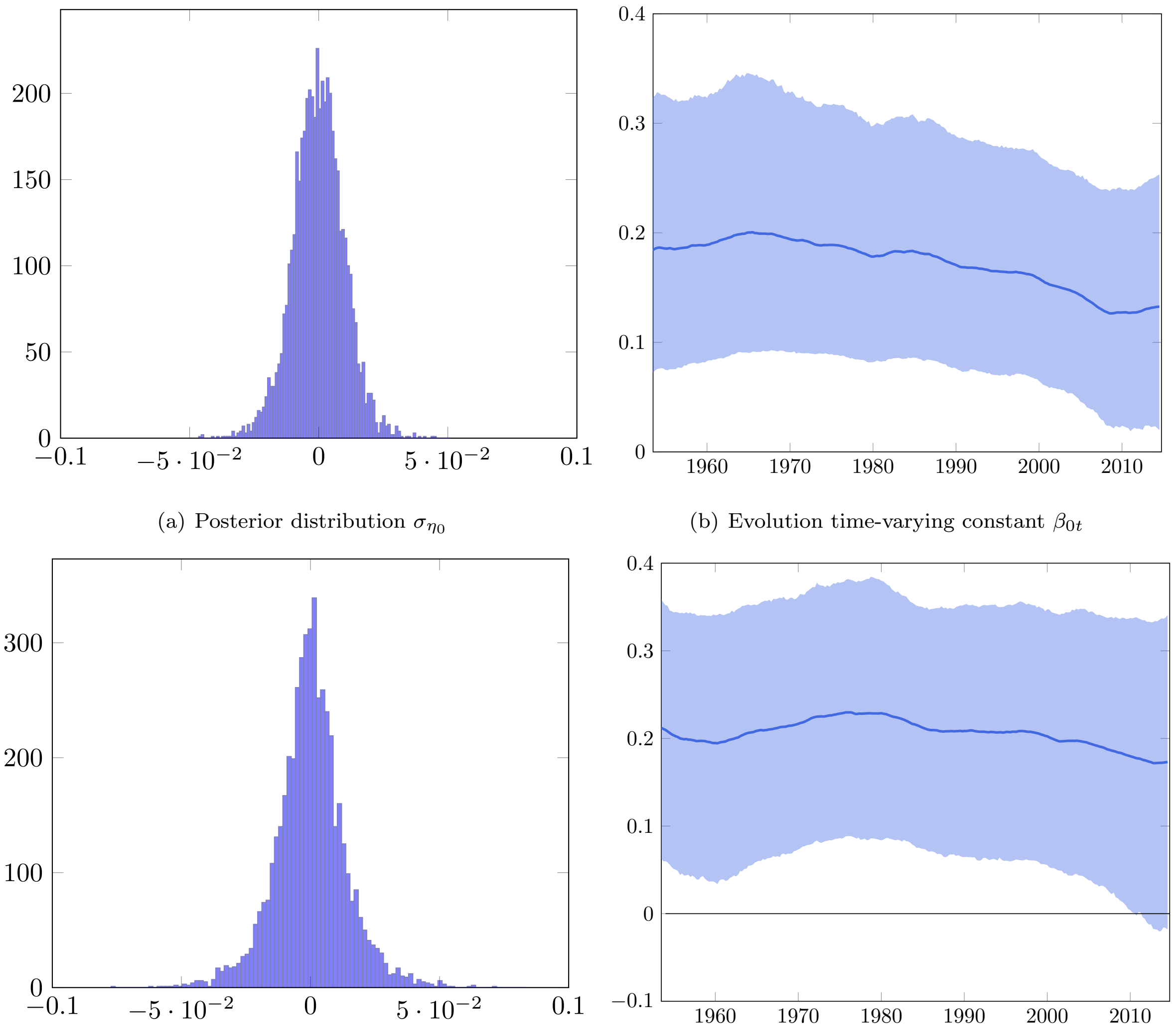
Panels (a) and (c) of Figure 4 clearly show that the posterior distributions of σ_{η_0} and σ_{η_1} are unimodal at zero. This suggests that these components are stable over time. Next, when sampling the stochastic binary indicators together with the other parameters, the results reported in Table 2 for M4 and M5 support these findings. The posterior probabilities for the binary indicators being one for the time-varying parts in the intercept and in the ES parameter are lower than 0.1 for all instrument sets. Even when increasing the prior inclusion probability to 0.9, the posterior probabilities of both indicators are not larger than 0.4. The model selection thus strongly rejects time variation in the intercept and in the ES of private consumption growth to expected disposable income growth. The unambiguous rejection of a time-varying intercept suggests that the omission of time-varying variables like hours worked and government consumption in our empirical specification is not a major source of concern. While, as M3 shows, there is still some indication of time variation in the intercept when lagged consumption growth is not included in the model, once we control for stickiness in aggregate consumption growth this is no longer the case. This confirms the results of Sommer (2007) and Carroll et al. (2011) who argue that allowing for consumption growth to depend on its own lag is important when testing for the ES of consumption to income. Further, when analyzing the MA structure of the residuals in Table 3, we notice that only the first MA term is relevant as the posterior mean of θ_1 equals -0.27 (in M4), respectively -0.22 (in M5) while the posterior distributions of θ_2 and - as noted above - θ_3 have considerable probability mass at zero.

Finally, as can be seen from Table 1, the prior beliefs about the degree of time variation in β_{0t} and β_{1t} are both centered at zero with a prior standard deviation of $\sqrt{V_0} = 0.2$. To check for robustness, we have also calculated the posterior inclusion probabilities of the time-varying components of β_{0t}^* and β_{1t}^*

[18] Note that it is the combination of habits and "classical" measurement error (which in itself leads to an MA(1) error) that leads to an MA(2) error. Likewise, it is the combination of habits and "general" measurement error (which in itself leads to an MA(2) error) that leads to an MA(3) error.

for M5 for alternative values for the prior standard deviation $\sqrt{V_0}$, i.e., for $\sqrt{V_0} = 0.05$ and $\sqrt{V_0} = 1$. We conclude - for all three prior specifications - that there is no evidence of time variation, neither in the intercept nor in the ES parameter. These results are not reported but are available from the authors upon request.

Figure 4: Stochastic model selection and time-varying parameters (binary indicators set to 1) in M5



(a) Posterior distribution $\sigma_{\eta 0}$

(b) Evolution time-varying constant β_{0t}

(c) Posterior distribution $\sigma_{\eta 1}$

(d) Evolution time-varying ES parameter β_{1t}

Note: Figures are presented for the results using instrument set Z^3 but are similar when using instrument sets Z^1 and Z^2

A parsimonious model

When allowing for stickiness in aggregate consumption growth, the time variation in both the intercept and the ES parameter is found to be irrelevant using the model selection criteria. We therefore restrict these parameters to be time invariant in the parsimonious model. Furthermore, the estimates of the MA parameters reported in Table 3 show that, for M4 and M5, only the first MA term is relevant. As such, we allow for only one MA term. This leads to the following parsimonious specification for aggregate consumption growth,

$$\Delta \ln C_t = \beta_{00} + \beta_{10} Z_t \delta + \gamma \Delta \ln C_{t-1} + \rho \nu_t^* + \theta(L) \mu_t,$$

with $\theta(L) = 1 + \theta_1 L$ an MA(1) lag polynomial.

Table 4: Posterior distributions of model parameters (parsimonious model)

			Percentiles	
		mean	2.5%	97.5%
error term consumption equation	σ_μ	0.39	0.36	0.43
error term income equation	σ_ν	0.80	0.73	0.87
			Percentiles	
Model parameters		mean	2.5%	97.5%
constant value of intercept	β_{00}	0.12	0.05	0.20
constant value of ES parameter	β_{10}	0.24	0.11	0.37
consumption stickiness	γ	0.55	0.41	0.70
degree of correlation between ν_t and ϵ_t	ρ	0.32	0.19	0.45
MA(1) parameter	θ_1	−0.30	−0.49	−0.10

Note: Results are presented using instrument set Z^3 but are similar when using instrument sets Z^1 and Z^2.

Descriptive statistics on the posterior distributions of the parsimonious model's parameters are given in Table 4. The results show that the time invariant intercept in the equation for aggregate consumption growth lies between 0.05 and 0.20.[19] For the ES parameter the 95% HPD interval varies between 0.11 and 0.37 with a mean value of 0.24. The stickiness parameter ranges between 0.41 and 0.70 with a mean of 0.55. Both these results are similar to findings reported by Carroll et al. (2011) and Kiley (2010)[20] and show that even if there is no significant amount of time variation, the ES of private consumption to disposable income remains a major factor contributing to the predictability of aggregate consumption growth. With respect to the presence of autocorrelation of the MA form in the error term, the posterior distribution points to a negative MA coefficient. The results also indicate that it is important to control for correlation between shocks to income growth and shocks to consumption growth as the 95% HPD interval of ρ ranges between 0.19 and 0.45.

[19] To interpret the magnitude of this estimate, note that data for aggregate consumption growth are expressed in percentage terms (as numbers between 0 and 100). The mean of aggregate consumption growth (on a quarterly basis) over the full sample period equals 0.48%.

[20] More specifically, for the US Carroll et al. (2011) find an ES parameter of 0.27 and a sticky consumption growth coefficient of 0.55 when using an instrumental variable approach. Kiley (2010) reports an ES parameter of 0.3 and a coefficient on lagged consumption growth of 0.65 when using their preferred instrument set (which includes lagged levels of inflation).

5 Conclusions

The recent literature investigates the excess sensitivity (ES) of aggregate consumption growth to anticipated aggregate disposable income growth using an elaborate empirical framework that contains both the possibility of stickiness in aggregate consumption growth and an adequate treatment of measurement error and time aggregation. However, this framework has only served as a benchmark for testing for ES under the assumption that the degree of ES is constant. This paper contributes to the literature by investigating *time-varying* ES in this elaborate empirical framework using quarterly US data over the period $1953-2014$. We estimate a Bayesian state space model using Markov Chain Monte Carlo (MCMC) methods. We test whether the time variation is statistically relevant using the Bayesian model selection approach recently suggested by Frühwirth-Schnatter and Wagner (2010). Their approach implies splitting the time-varying ES parameter, which is assumed to follow a standard random walk process, into a constant part and into a time-varying part and introducing a stochastic binary model indicator which is one if the time-varying part should be included in the model and zero otherwise. To control for endogeneity in our framework, we further incorporate a control function type approach to instrumental variables estimation in our MCMC algorithm. As our Bayesian IV approach relies on sampling the posterior distribution rather than using asymptotic approximations, it allows for exact inference even when instruments are weak.

The estimation results show that in a basic model that includes only anticipated income growth, the ES of US consumption growth to anticipated income growth has decreased over time, starting from around 0.4 in the early 1950s and ending close to 0.25 in 2014. This confirms some of the results reported in the literature that argue that excess sensitivity has dropped gradually over time. However, when estimating our elaborate empirical specification that includes the possibility of stickiness in consumption growth along with the possibility of time aggregation and measurement error, the excess sensitivity parameter is found to be stable at around 0.24 over the entire sample period. This suggests that the time variation of the ES parameter found in the basic model is due to a specification error. In line with Carroll et al. (2011) and others, the coefficient on lagged consumption growth is found to be around 0.55 showing that there is a notable amount of stickiness in aggregate consumption growth.

6 Acknowledgements

The authors would like to thank Tino Berger, Markus Eberhardt, Freddy Heylen and Glenn Rayp for their constructive comments and suggestions. The paper has also benefited from discussions with participants at the 23rd Symposium of the Society for Nonlinear Dynamics and Econometrics and various seminars

and workshops.

References

Bacchetta, P. and Gerlach, S. (1997). Consumption and credit constraints: international evidence. *Journal of Monetary Economics*, 40:207–38.

Basu, S. and Kimball, M. (2002). Long-run labor supply and the elasticity of intertemporal substitution for consumption. Mimeo, University of Michigan.

Blundell-Wignall, A., Browne, F., and Cavaglia, S. (1991). Financial liberalization and consumption behaviour. *OECD Working Papers*, 81.

Campbell, J. (1987). Does saving anticipate declining labor income - An alternative test of the ppermanent income hypothesis. *Econometrica*, 55(6):1249–1273.

Campbell, J. and Mankiw, N. (1989). Consumption, income, and interest rates: reinterpreting the time series evidence. *NBER Macroeconomics Annual*, 4:185–216.

Campbell, J. and Mankiw, N. (1990). Permanent income, current income and consumption. *Journal of Business and Economic Statistics*, 8:265–79.

Campbell, J. and Mankiw, N. (1991). The response of consumption to income: a cross-country investigation. *European Economic Review*, 35:723–67.

Carroll, C. (1992). The buffer-stock theory of saving: some macroeconomic evidence. *Brookings Papers on Economic Activity*, 2:61–156.

Carroll, C., Slacalek, J., and Sommer, M. (2011). International evidence on sticky consumption growth. *Review of Economics and Statistics*, 93:1135–1145.

Carter, C. and Kohn, R. (1994). On gibbs sampling for state space models. *Biometrika*, 81:541–53.

Chib, S. and Greenberg, E. (1994). Bayes inference in regression models with ARMA(p,q) errors. *Journal of Econometrics*, 64:183–206.

De Jong, P. and Shephard, N. (1995). The simulation smoother fro time series models. *Biometrika*, 82:339–50.

Deaton, A. (1991). Savings and liquidity constraints. *Econometrica*, 59:1,221–48.

Dynan, K. (2000). Habit formation in consumer preferences: evidence from panel data. *American Economic Review*, 90(3):391–406.

Evans, P. and Karras, G. (1998). Liquidity constraints and the substitutability between private and government consumption: the role of military and non-military spending. *Economic Inquiry*, 36:203–14.

Flavin, M. (1985). Excess sensitivity of consumption to current income: liquidity constraints or myopia? *Canadian Journal of Economics*, 38:117–36.

Frühwirth-Schnatter, S. and Wagner, H. (2010). Stochastic model specification search for Gaussian and partial non-Gaussian state space models. *Journal of Econometrics*, 154(1):85–100.

Fuhrer, J. (2000). Habit formation in consumption and its implications for monetary-policy models. *American Economic Review*, 90(3):367–390.

Gali, J., Lopez-Salido, D., and Valles, J. (2007). Understanding the effects of government spending on consumption. *Journal of the European Economic Association*, 5:227–270. DOI: 10.1162/JEEA.2007.5.1.227.

George, E. and McCulloch, R. (1993). Variable selection via gibbs sampling. *Journal of the American Statistical Association*, 88:881–889.

Girardin, E., Sarno, L., and Taylor, M. (2000). Private consumption behaviour, liquidity constraints and financial deregulation in france: a nonlinear analysis. *Empirical Economics*, 25:351–368.

Hall, R. (1978). Stochastic implications of the life cycle-permanent income hypothesis: theory and evidence. *Journal of Political Economy*, 86:971–87.

Hamilton, J. (1994). *Time series analysis*. Princeton.

Hayashi, F. (1985). The permanent income hypothesis and consumption durability: analysis based on Japanese panel data. *Quarterly Journal of Economics*, 100(4):1083–1113.

Kiley, M. (2010). Habit persistence, nonseparability between consumption and leisure, or rule-of-thumb consumers: which accounts for the predictability of consumption growth? *Review of Economics and Statistics*, 92:679–683.

Kim, Y. and Kim, C.-J. (2011). Dealing with endogeneity in a time-varying parameter model: joint estimation and two-step estimation procedures. *Econometrics Journal*, 14:487–497.

Ludvigson, S. (1999). Consumption and credit: a model of time-varying liquidity constraints. *Review of Economics and Statistics*, 81(3):434–47.

Ludvigson, S. and Michaelides, A. (2001). Does buffer-stock saving explain the smoothness and excess sensitivity of consumption. *American Economic Review*, 91(3):631–47.

McKiernan, B. (1996). Consumption and the credit market. *Economics Letters*, 51:83–88.

Muellbauer, J. (1988). Habits, rationality and myopia in the life cycle consumption function. *Annales d'Economie et de Statistique*, 9:47–70.

Peersman, G. and Pozzi, L. (2007). Business cycle fluctuations and excess sensitivity of private consumption. *Economica*, 75:1–10.

Pischke, J. (1995). Individual income, incomplete information and aggregate consumption. *Econometrica*, 63:805–40.

Reis, R. (2006). Inattentive consumers. *Journal of Monetary Economics*, 53:1761–1800.

Sommer, M. (2007). Habit formation and aggregate consumption dynamics. *The B.E. Journal of Macroeconomics - Advances*, 7:article 21.

Ullah, A., Vinod, H., and Singh, R. (1986). Estimation of linear models with moving average disturbances. *Journal of Quantitative Economics*, 2:137–152.

Appendix A Gibbs sampling algorithm

In this appendix we provide details on the Gibbs sampling algorithm used in section 3.3 to jointly sample the binary indicators $\mathcal{M}$, the parameters ϕ and the time-varying parameters β^*.

Blocks 1-3: Sampling the binary indicators $\mathcal{M}$ and the parameters ϕ

For notational convenience, let us define a general regression model

$$y_t = x_t^{\mathcal{M}} b^{\mathcal{M}} + \theta(L) e_t, \qquad e_t \sim \mathcal{N}\left(0, \sigma_e^2\right), \tag{B-1}$$

where y_t is a scalar dependent variable, x_t an unrestricted predictor vector that contains variables that are relevant for explaining y_t, b is the corresponding parameter vector, $\theta(L)$ is a lag polynomial of order q and e_t is a white noise error with variance σ_e^2. The restricted predictor matrix $x_t^{\mathcal{M}}$ and restricted parameter vector $b^{\mathcal{M}}$ exclude those elements in x_t and b for which the corresponding binary indicator in

$\mathcal{M}$ is 0. Further let $y = [y_1, \ldots, y_T]'$, $x = [x_1', \ldots, x_T']'$ and Φ be a subset of ϕ including all unknown parameters in equation (B-1), with restricted versions $x^{\mathcal{M}}$ and $\Phi^{\mathcal{M}}$.

The MA(q) errors in equation (B-1) imply a model which is non-linear in the parameters. As suggested by Ullah et al. (1986) and Chib and Greenberg (1994), conditional on θ a linear model can be obtained from a recursive transformation of the data. For $t = 1, \ldots T$ let

$$\widetilde{y}_t = y_t - \sum_{i=1}^{q} \theta_i \widetilde{y}_{t-i}, \qquad \text{with} \quad \widetilde{y}_t = 0 \quad \text{for } t \leq 0, \tag{B-2}$$

$$\widetilde{x}_t = x_t - \sum_{i=1}^{q} \theta_i \widetilde{x}_{t-i}, \qquad \text{with} \quad \widetilde{x}_t = 0 \quad \text{for } t \leq 0, \tag{B-3}$$

and further for $j = 1, \ldots, q$

$$\omega_{jt} = -\sum_{i=1}^{q} \theta_i \omega_{j,t-i} + \theta_{t+j-1}, \qquad \text{with} \quad \omega_{jt} = 0 \quad \text{for } t \leq 0, \tag{B-4}$$

where $\theta_s = 0$ for $s > q$. Equation (B-1) can then be transformed as

$$\begin{aligned} \widetilde{y}_t &= \widetilde{x}_t^{\mathcal{M}} b^{\mathcal{M}} + \omega_t \lambda + e_t, \\ &= \widetilde{w}_t^{\mathcal{M}} \Phi^{\mathcal{M}} + e_t, \end{aligned} \tag{B-5}$$

with $\omega_t = (\omega_{1t}, \ldots, \omega_{qt})$, $\widetilde{w}_t = (\widetilde{x}_t, \omega_t)$ and $\Phi^{\mathcal{M}} = \left(b^{\mathcal{M}'}, \lambda'\right)'$ and where $\lambda = (e_0, \ldots, e_{-q+1})'$ are initial conditions that can be estimated as unknown parameters.

Conditional on θ, equation (B-5) is a standard linear regression with observed variables $\widetilde{y}_t$ and $\widetilde{w}_t^{\mathcal{M}}$ and *i.i.d.* errors e_t. Under the normal-inverse gamma conjugate prior[21]

$$p\left(\Phi^{\mathcal{M}}\right) = \mathcal{N}\left(b_0^{\mathcal{M}}, B_0^{\mathcal{M}} \sigma_e^2\right), \qquad p\left(\sigma_e^2\right) = \mathcal{IG}\left(c_0, C_0\right), \tag{B-6}$$

the conditional posterior distributions of $\Phi^{\mathcal{M}}$ and σ_e^2 are

$$p\left(\Phi^{\mathcal{M}} | y, x, \theta, \sigma_e^2, \mathcal{M}\right) = \mathcal{N}\left(b_T^{\mathcal{M}}, B_T^{\mathcal{M}} \sigma_e^2\right), \qquad p\left(\sigma_e^2 | y, x, \theta, \mathcal{M}\right) = \mathcal{IG}\left(c_T, C_T^{\mathcal{M}}\right), \tag{B-7}$$

[21]Note that we set prior variances V_0 in Table 1 from which B_0 can be calculated as $B_0 = V_0/\sigma_0^2$ with σ_0^2 the prior variance of the error terms in either the consumption or income growth equation.

with the posterior moments $b_T^{\mathcal{M}}$, $B_T^{\mathcal{M}}$, c_T and $C_T^{\mathcal{M}}$ given by

$$b_T^{\mathcal{M}} = B_T^{\mathcal{M}} \left(\left(\widetilde{w}^{\mathcal{M}}\right)' \widetilde{y} + \left(B_0^{\mathcal{M}}\right)^{-1} b_0^{\mathcal{M}} \right), \tag{B-8}$$

$$B_T^{\mathcal{M}} = \left(\left(\widetilde{w}^{\mathcal{M}}\right)' \widetilde{w}^{\mathcal{M}} + \left(B_0^{\mathcal{M}}\right)^{-1} \right)^{-1}, \tag{B-9}$$

$$c_T = c_0 + T/2 , \tag{B-10}$$

$$C_T^{\mathcal{M}} = C_0 + 0.5 \left(\widetilde{y}'\widetilde{y} + \left(b_0^{\mathcal{M}}\right)' \left(B_0^{\mathcal{M}}\right)^{-1} b_0^{\mathcal{M}} - \left(b_T^{\mathcal{M}}\right)' \left(B_T^{\mathcal{M}}\right)^{-1} b_T^{\mathcal{M}} \right). \tag{B-11}$$

Block 1: Sampling the first step parameters ϕ_1 and calculating $Z_t\delta$ and ν_t^*

Equation (14) can be written in the general notation of equation (B-1) as: $y_t = \Delta \ln Y_t$, $x_t = Z_t$, $b = \delta$ and $\theta(L) = 1$ such that $\theta(L)e_t = \nu_t$ and $\sigma_e^2 = \sigma_\nu^2$. Sampling δ and σ_ν^2 can then be done from their posterior distributions in equation (B-7). Using the sampled δ and σ_ν^2, calculate $E_{t-1}(\Delta \ln Y_t) = Z_t\delta$ and $\nu_t^* = \sigma_\varepsilon \theta(L)(\Delta \ln Y_t - Z_t\delta)/\sigma_\nu$ conditional on θ and σ_ε^2 with the latter calculated from ϕ_2 as $\sigma_\varepsilon^2 = \sigma_\mu^2 / (1-\rho^2)$.

Block 2: Sampling the MA coefficients θ

Conditional on the parameters ϕ_1 and ϕ_2, on the time-varying coefficients β^* and on the binary indicators $\mathcal{M}$, equation (16) can be written in the general notation of equation (B-1) as: $y_t = \Delta \ln C_t$, $x_t = (1,$ $Z_t\delta$, β_{0t}^*, $\beta_{1t}^* Z_t\delta$, $\Delta \ln C_{t-1})$, $b = (\beta_{00}, \beta_{10}, \sigma_{\eta_0}, \sigma_{\eta_1}, \gamma)$ and $e_t = \varepsilon_t$, such that $\sigma_e^2 = \sigma_\varepsilon^2$ with the latter calculated conditional on ϕ_2 as $\sigma_\varepsilon^2 = \sigma_\mu^2 / (1-\rho^2)$. The values of the binary indicators in $\mathcal{M}$ then imply the restricted $x_t^{\mathcal{M}}$ and $b^{\mathcal{M}}$.

Under the normal conjugate prior $p(\theta) = \mathcal{N}\left(b_0^\theta, B_0^\theta \sigma_e^2\right)$, the exact conditional distribution of θ is given by[22]

$$p\left(\theta | \Phi, \sigma_e^2, \mathcal{M}, y, x\right) \propto \prod_{t=1}^{T} \exp\left(-\frac{e_t(\theta)^2}{2\sigma_e^2} \right) \times \exp\left(-\frac{1}{2} \left(\theta - b_0^\theta\right)' \left(B_0^\theta \sigma_e^2\right)^{-1} \left(\theta - b_0^\theta\right) \right), \tag{B-12}$$

where $e_t(\theta) = \widetilde{y}_t(\theta) - \widetilde{w}_t^{\mathcal{M}}(\theta)\Phi^{\mathcal{M}}$ is calculated from the transformed model in equation (B-5) further conditioning on the initial conditions λ to obtain $\Phi^{\mathcal{M}} = \left(b^{\mathcal{M}\prime}, \lambda'\right)'$.

Direct sampling of θ using equation (B-12) is not possible, though, as $e_t(\theta)$ is a non-linear function of θ. To solve this issue, Chib and Greenberg (1994) propose to linearize $e_t(\theta)$ around θ^* using a first-order

[22] Note that the expression in Chib and Greenberg (1994) also includes a term ($p_2(\theta)$ in their notation) which evaluates the initial conditions (α_0 in their notation) which are drawn (using a value for θ) as initial values in a state space representation. As is apparent from equation (B-5), in a pure MA model (see also Chib and Greenberg, 1994, eq. (15)), the initial conditions are easily estimated together with β. As such, they are conditioned on in equation (B-12).

Taylor expansion

$$e_t(\theta) \approx e_t(\theta^*) - \Psi_t(\theta - \theta^*), \tag{B-13}$$

where $\Psi_t = (\Psi_{1t}, \ldots, \Psi_{qt})$ is a $1 \times q$ vector including the first-order derivatives of $e_t(\theta)$ evaluated at θ^* obtained using the following recursion

$$\Psi_{it} = -e_{t-i}(\theta^*) - \sum_{j=1}^{q} \theta_j^* \Psi_{i,t-j}, \tag{B-14}$$

where $\Psi_{it} = 0$ for $t \leq 0$. An adequate approximation can be obtained by choosing θ^* to be the non-linear least squares estimate of θ conditional on the other parameters in the model, which can be obtained as

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \sum_{t=1}^{T} (e_t(\theta))^2, \tag{B-15}$$

For given values of θ^*, equation (B-13) can then be rewritten as an approximate linear regression model

$$e_t(\theta^*) + \Psi_t \theta^* \approx \Psi_t \theta + e_t(\theta), \tag{B-16}$$

with dependent variable $e_t(\theta^*) + \Psi_t \theta^*$ and explanatory variables Ψ_t. As such a normal approximation to the exact conditional distribution of θ is given by

$$q\left(\theta | \theta^*, \Phi, \sigma_e^2, \mathcal{M}, y, x\right) \sim \mathcal{N}\left(b_T^\theta, B_T^\theta \sigma_e^2\right), \tag{B-17}$$

with

$$b_T^\theta = B_T^\theta \left(\Psi' \Xi + \left(B_0^\theta\right)^{-1} b_0^\theta\right), \qquad B_T^\theta = \left(\Psi'\Psi + \left(B_0^\theta\right)^{-1}\right)^{-1}, \tag{B-18}$$

and where Ξ is a $T \times 1$ vector with tth element $(e_t(\theta^*) + \Psi_t \theta^*)$ and Ψ is a $T \times q$ matrix with tth row Ψ_t.

We can now sample θ using a Metropolis-Hastings (MH) algorithm. Suppose $\theta^{(i)}$ is the current draw in the Markov chain. To obtain the next draw $\theta^{(i+1)}$, first draw a candidate θ^c from the proposal distribution in equation (B-17). The MH step then implies a further randomization which amounts to accepting the candidate draw θ^c with probability

$$\alpha\left(\theta^{(i)}, \theta^c\right) = \min\left\{\frac{p\left(\theta^c | \Phi, \sigma_e^2, \mathcal{M}, y, x\right)}{p\left(\theta^{(i)} | \Phi, \sigma_e^2, \mathcal{M}, y, x\right)} \frac{q\left(\theta^{(i)} | \theta^*, \Phi, \sigma_e^2, \mathcal{M}, y, x\right)}{q\left(\theta^c | \theta^*, \Phi, \sigma_e^2, \mathcal{M}, y, x\right)}, 1\right\}. \tag{B-19}$$

If θ^c is accepted, $\theta^{(i+1)}$ is set equal to θ^c while if θ^c is rejected, $\theta^{(i+1)}$ is set equal to $\theta^{(i)}$.

Block 3: Sampling the binary indicators $\mathcal{M}$ and the second step parameters ϕ_2

Conditional on the time-varying coefficients β_t^* and on the first block results $Z_t\delta$ and ν_t^*, equation (24) can be written in the general notation of equation (B-1) as: $y_t = \Delta \ln C_t$, $x_t = (1,\ Z_t\delta,\ \beta_{0t}^*,\ \beta_{1t}^* Z_t\delta,$ $\Delta \ln C_{t-1},\ \nu_t^*)$, $b = (\beta_{00},\ \beta_{01},\ \sigma_{\eta_0},\ \sigma_{\eta_1},\ \gamma,\ \rho)$ and $e_t = \mu_t$, such that $\sigma_e^2 = \sigma_\mu^2$. Further conditioning on the MA parameters θ, the unrestricted transformed variables $\widetilde{y}_t$ and $\widetilde{w}_t$ in equation (B-5) are obtained, with corresponding unrestricted extended parameter vector $\Phi = (b', \lambda')'$. The values of the binary indicators in $\mathcal{M}$ then imply the restricted $\widetilde{w}_t^{\mathcal{M}}$ and $\Phi^{\mathcal{M}}$.

A naive implementation of the Gibbs sampler would be to first sample $\mathcal{M}$ from $f\left(\mathcal{M}|\Phi, \sigma_e^2, \widetilde{y}, \widetilde{w}\right)$ and next $\Phi^{\mathcal{M}}$ and σ_e^2 from $f\left(\Phi^{\mathcal{M}}, \sigma_e^2|\mathcal{M}, \widetilde{y}, \widetilde{w}\right)$. However, this approach does not result in an irreducible Markov chain as whenever an indicator in $\mathcal{M}$ equals zero, the corresponding coefficient in Φ is also zero which implies that the chain has absorbing states. Therefore, as in Frühwirth-Schnatter and Wagner (2010) we marginalize over the parameters Φ and σ_e^2 when sampling $\mathcal{M}$ and next draw the parameters $\Phi^{\mathcal{M}}$ and σ_e^2 conditional on the binary indicators in $\mathcal{M}$.

Block 3(a): Sampling the binary indicators $\mathcal{M}$

The posterior distribution $f(\mathcal{M}|\widetilde{y}, \widetilde{w})$ can be obtained using Bayes' Theorem as

$$f(\mathcal{M}|\widetilde{y}, \widetilde{w}) \propto f(\widetilde{y}\,|\mathcal{M}, \widetilde{w}\,)\, p(\mathcal{M}), \tag{B-20}$$

with $p(\mathcal{M})$ being the prior probability of $\mathcal{M}$ and $f(\widetilde{y}\,|\mathcal{M}, \widetilde{w}\,)$ the marginal likelihood of the regression model (B-5) where the effect of the parameters Φ and σ_e^2 has been integrated out. Under the normal-inverse gamma conjugate prior in equation (B-6), the closed form solution of the marginal likelihood is given by:

$$f(\widetilde{y}\,|\mathcal{M}, \widetilde{w}\,) \propto \frac{\left|B_T^{\mathcal{M}}\right|^{0.5}}{\left|B_0^{\mathcal{M}}\right|^{0.5}} \frac{\Gamma(c_T)\, C_0^{c_0}}{\Gamma(c_0)\left(C_T^{\mathcal{M}}\right)^{c_T}}, \tag{B-21}$$

with Γ being the gamma function and the posterior moments $b_T^{\mathcal{M}}$, $B_T^{\mathcal{M}}$, c_T and $C_T^{\mathcal{M}}$ given in equations (B-8)-(B-11).

Following George and McCulloch (1993) we use a single-move sampler in which the binary indicators ι_0 and ι_1 in $\mathcal{M}$ are sampled recursively from the Bernoulli distribution with probability

$$p(\iota_i = 1\,|\iota_{-i}, \widetilde{y}, \widetilde{w}\,) = \frac{f(\iota_i = 1\,|\iota_{-i}, \widetilde{y}, \widetilde{w}\,)}{f(\iota_i = 0\,|\iota_{-i}, \widetilde{y}, \widetilde{w}\,) + f(\iota_i = 1\,|\iota_{-i}, \widetilde{y}, \widetilde{w}\,)}, \tag{B-22}$$

for $i = 0, 1$. We further randomize over the sequence in which the binary indicators are drawn.

Block 3(b): Sampling the second step parameters ϕ_2

Given the binary indicators in $\mathcal{M}$, the second step parameters $\phi_2 = (\beta_{00},\ \beta_{10},\ \sigma_{\eta_0},\ \sigma_{\eta_1},\ \gamma,\ \rho,\ \sigma_\mu^2)$ are sampled, together with λ, by drawing $\Phi^{\mathcal{M}}$ and σ_e^2 from the general expression in equation (B-7). Note that the unrestricted $\Phi = (\beta_{00},\ \beta_{10},\ \sigma_{\eta_0},\ \sigma_{\eta_1},\ \gamma,\ \rho,\ \lambda)$ is restricted to obtain $\Phi^{\mathcal{M}}$ by excluding σ_{η_i} when $\iota_i = 0$. In this case σ_{η_i} is not sampled but set equal to zero.

Block 4: Sampling the time-varying parameters β^*

In this block we use the forward-filtering and backward-sampling approach of Carter and Kohn (1994) and De Jong and Shephard (1995) to sample the time-varying parameters β^* conditionally on the coefficients ϕ_2 and λ, on the first block results $Z_t\delta$ and ν_t^* and on the binary indicators $\mathcal{M}$. More specifically, equation (24) can be rewritten as:

$$y_t = \iota_0\sigma_{\eta_0}\beta_{0t}^* + \iota_1\sigma_{\eta_1}\beta_{1t}^* x_{1t} + \theta\left(L\right)\mu_t, \tag{B-23}$$

with $y_t = \Delta\ln C_t - \beta_{00} - \beta_{10}Z_t\delta - \gamma\Delta\ln C_{t-1} - \rho\nu_t^*$ and $x_{1t} = Z_t\delta$.

Again using the recursive transformation suggested by Ullah et al. (1986) and Chib and Greenberg (1994), the model in equation (B-23) can be transformed to a model with *i.i.d.* error terms as

$$\widetilde{y}_t = \iota_0\sigma_{\eta_0}\widetilde{\beta}_{0t} + \iota_1\sigma_{\eta_1}\widetilde{\beta}_{1t} + \omega_t\lambda + \mu_t, \tag{B-24}$$

where $\widetilde{y}_t$ and $\omega_t = (\omega_{1t}, \ldots, \omega_{qt})$ are calculated (conditional on θ) from equations (B-2) and (B-4) and similarly

$$\widetilde{\beta}_{0t} = \beta_{0t}^* - \sum_{i=1}^{q}\theta_i\widetilde{\beta}_{0,t-i}, \qquad \text{with} \quad \widetilde{\beta}_{0t} = 0 \quad \text{for } t \leq 0, \tag{B-25}$$

$$\widetilde{\beta}_{1t} = \beta_{1t}^* x_{1t} - \sum_{i=1}^{q}\theta_i\widetilde{\beta}_{1,t-i}, \qquad \text{with} \quad \widetilde{\beta}_{1t} = 0 \quad \text{for } t \leq 0. \tag{B-26}$$

Substituting equation (22) in (B-25)-(B-26) yields

$$\widetilde{\beta}_{0,t+1} = \beta_{0t}^* - \sum_{i=1}^{q}\theta_i\widetilde{\beta}_{0,t+1-i} + \eta_{0t}^*, \tag{B-27}$$

$$\widetilde{\beta}_{1,t+1} = \beta_{1t}^* x_{1,t+1} - \sum_{i=1}^{q}\theta_i\widetilde{\beta}_{1,t+1-i} + x_{1,t+1}\eta_{1t}^*, \tag{B-28}$$

such that the state space representation of the model in equations (B-24), (22) and (B-27)-(B-28) is given by

$$\widetilde{y}_t - \omega_t \lambda = \overbrace{\left[\begin{array}{cc} (0 \quad \sigma_{\eta_0} \quad 0 \quad \ldots \quad 0) & (0 \quad \sigma_{\eta_1} \quad 0 \quad \ldots \quad 0) \end{array}\right]}^{\sigma_\eta} \overbrace{\left[\begin{array}{c} \alpha_{0t} \\ \alpha_{1t} \end{array}\right]}^{\alpha_t} + \mu_t, \tag{B-29}$$

$$\underbrace{\left[\begin{array}{c} \alpha_{0,t+1} \\ \alpha_{1,t+1} \end{array}\right]}_{\alpha_{t+1}} = \underbrace{\left[\begin{array}{cc} T_{0t} & 0 \\ 0 & T_{1t} \end{array}\right]}_{T_t} \underbrace{\left[\begin{array}{c} \alpha_{0t} \\ \alpha_{1t} \end{array}\right]}_{\alpha_t} + \underbrace{\left[\begin{array}{cc} K_{0t} & 0 \\ 0 & K_{1t} \end{array}\right]}_{K_t} \underbrace{\left[\begin{array}{c} \eta^*_{0t} \\ \eta^*_{1t} \end{array}\right]}_{\eta_t}, \tag{B-30}$$

with $\alpha_{i,t+1}$ given by

$$\underbrace{\left[\begin{array}{c} \beta^*_{i,t+1} \\ \widetilde{\beta}_{i,t+1} \\ \widetilde{\beta}_{it} \\ \vdots \\ \widetilde{\beta}_{i,t-(q-2)} \end{array}\right]}_{\alpha_{i,t+1}} = \underbrace{\left[\begin{array}{ccccc} 1 & 0 & \ldots & 0 & 0 \\ x_{i,t+1} & -\theta_1 & \ldots & -\theta_{q-1} & -\theta_q \\ 0 & 1 & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{array}\right]}_{T_{it}} \underbrace{\left[\begin{array}{c} \beta^*_{it} \\ \widetilde{\beta}_{it} \\ \vdots \\ \widetilde{\beta}_{i,t-(q-2)} \\ \widetilde{\beta}_{i,t-(q-1)} \end{array}\right]}_{\alpha_{it}} + \underbrace{\left[\begin{array}{c} 1 \\ x_{i,t+1} \\ 0 \\ \vdots \\ 0 \end{array}\right]}_{K_{it}} \left[\begin{array}{c} \eta^*_{it} \end{array}\right], \tag{B-31}$$

for $i = 0, 1$ and where $x_{0t} = 1 \; \forall t$. In line with equations (22) and (B-25)-(B-26), each of the states is initialized at zero.

Equations (B-29)-(B-30) constitute a standard linear Gaussian state space model, from which the unknown state variables α_t can be filtered using the standard Kalman filter. Sampling α_t from its conditional distribution can then be done using the multimove simulation smoother of Carter and Kohn (1994) and De Jong and Shephard (1995). Using β_{i0}, σ_{η_i} and β^*_{it}, the time-varying coefficients β_{it} in equation (20) can then easily be reconstructed from equation (21). Note that in a restricted model with $\iota_i = 0$, σ_{η_i} is excluded from σ_η and α_{it} is dropped from the state vector α_t. In this case, no forward-filtering and backward-sampling for β^*_{it} is needed as this can be sampled directly from its prior distribution using equation (22).