On-line Appendices
Dealing with misspecification in structural macroeconometric models

Fabio Canova, Norwegian Business School *

Christian Matthes, Indiana University †

October 19, 2020

*Norwegian Business School, CAMP, and CEPR. Department of Economics, BI Norwegian Business School, Nydalsveien 37, 0484 Oslo, Norway; email: fabio.canova@bi.no

†Department of Economics, 100 S Woodlawn Avenue, Bloomington, IN 47405, USA. email: matthesc@iu.edu.

Appendix A: Classical composite likelihood estimators

Asymptotic properties of composite likelihood estimators In a standard approach one has a known DGP which produces a parametric density $F(y_t, \psi)$ for an $m \times 1$ vector of observables y_t, given a $q \times 1$ vector of parameters $\psi = (\theta, \eta)$, where θ is $q_1 \times 1$ and η is $q - q_1 \times 1$. When y_t is of high dimensions or contains latent variables, it may be difficult to use $F(y_t, \psi)$ for estimation.

The key idea of composite methods is to construct arbitrary sets of low dimensional densities and to combine them for estimation purposes. This may be viewed as a *divide-and-conquer* method of approximating the full likelihood.

Let $f(y_{it} \in A_i, \phi_i)$ be sub-densities of $F(y_t, \psi)$ obtained by marginalizing (or conditioning on portions of) $F(y_t, \psi)$, where A_i is a set and $i = 1, \ldots, K$. For ease of reading, the integrals and the conditioning sets are left implicit. Each sub-density defines a sub-model, has an associated vector of parameters $\phi_i = [\theta, \eta_i]'$, where η_i are (nuisance) sub-density specific, and has implications for a sub-vector y_{it} of length T_i. The elements of y_{it} need not be mutually exclusive across i and T_i may be different than T_j. Given a vector of fixed weights ω_i, the composite likelihood is

$$CL(\theta, \eta_1, \ldots, \eta_K, y_{1t}, \ldots, y_{KT}) = \Pi_{i=1}^{K} f(y_{it} \in A_i, \theta, \eta_i)^{\omega_i} \equiv \Pi_{i=1}^{K} \mathcal{L}(\theta, \eta_i | y_{it} \in A_i)^{\omega_i} \tag{1}$$

Although $CL(\phi, y) \equiv CL(\theta, \eta_1, \ldots, \eta_K, y_{1t}, \ldots, y_{KT})$ is not a likelihood function, if $y_{[1,t]} = (y_1, \ldots, y_t)$ is an independent sample from $F(y_t, \psi)$ and ω_i are fixed, θ_{CL}, the maximum composite likelihood estimator satisfies $\theta_{CL} \xrightarrow{P} \theta$ and

$$\sqrt{T}(\theta_{CL} - \theta) \xrightarrow{D} N(0, G^{-1}) \tag{2}$$

for T going to infinity, K fixed (see e.g. Varin, et al., 2011) where

$$\begin{aligned}
G &= HJ^{-1}H \text{ Godambe information} && (3)\\
J &\equiv var_\theta u(\phi, y_{[1,t]}|\omega) \quad \text{Variability matrix} && (4)\\
H &\equiv -E_\theta[\bigtriangledown_\theta u(\phi, y_{[1,t]}|\omega)] \quad \text{Sensitivity matrix} && (5)\\
u(\phi, y_{[1,t]}|\omega) &= \sum_i \omega_i \bigtriangledown_\theta\, l_i(\theta, \eta_i, y_{[1,t]}) \quad \text{Composite scores} && (6)
\end{aligned}$$

and $\bigtriangledown_\theta l_i(\theta, \eta_i, y_{[1,t]})$ are the scores associated with the log of $f(y_{it} \in A_i, \theta, \eta_i)$, and $H \neq J$ [1].

Consistency obtains because each sub-model i provides an unbiased estimating function for θ. Since the ML estimator of each sub-model converges to the true parameter vector as T increases, $\theta_{CL} \xrightarrow{P} \theta$. Asymptotic normality holds because the sampling distribution of the maximum likelihood estimator of each i can be approximated quadratically around the same mode. θ_{CL} is

[1] If T is fixed, but $K \to \infty$, and the sub-models are independent, the result still holds. On the other hand, when $\{y_t\}_{t=1}^{T}$ has correlated observations similar results can be proved, see Engle et al. (2008). Note also that a standard Newey-West correction to $J(\theta)$ can be used if $y_{[1,t]}$ is not an independent sample.

inefficient - G equals Fisher information matrix, I, only if the composite likelihood is the likelihood of the true model. Careful choices ω_i may improve efficiency and optimal weights can be designed by minimizing the distance between G and I, or by insuring that the composite likelihood ratio statistics has an asymptotic χ^2 distribution, see Pauli et al. (2011).

If consistency is all that one cares about, one could set $\omega_i = \frac{1}{K}, \forall i$ or use a data-based approach, e.g. select $\omega_i = \frac{\exp(\zeta_i)}{1+\sum_{i=1}^{K-1}\exp(\zeta_i)}$, where ζ_i are functions of some statistics of past data, $\zeta_i = \zeta(Y_{i,[-\tau:0]})$. If these statistics are updated over time, ω_i could also be made time varying. There is a large forecasting literature (see e.g. Aiolfi et al., 2010) which can be used to select training sample-based estimates of ω_i and to make them time varying.

The asymptotic properties of θ depend on $(\eta_1, \ldots, \eta_K)$. In standard exercises η_i are assumed to be known, so the dependence disappears. When η_i are unknown, but estimable a two-step approach is generally implemented: η_i are estimated from each $\log f(y_{it} \in A_i, \theta, \eta_i)$ and plugged in the composite likelihood, which is then optimized with respect to θ, see e.g. Pakel et al. (2011). Consistency of θ_{CL} is unaffected as long as η_i are consistently estimated, but standard errors need to be properly adjusted. A two-step approach is convenient when K or the number of nuisance parameters is large, since joint estimation of $(\theta, \eta_1, \ldots, \eta_K)$ may be demanding.

Asymptotic properties of composite estimators under misspecification When $f(y_{it} \in A, \theta, \eta_i)$ are not marginal or conditional representations of $F(y_t, \psi)$, the previous conclusions need to be modified. Let $y_{[1,t]}$ be a sample from $F(y_t, \psi)$ with respect to some σ-measure μ. Suppose model i with density $f_i(y_{[1,t]}, \phi_i)$, where $\phi_i \in \Phi \subset R^m$ is a vector of parameters, is used in the analysis and let its log-likelihood be $l_i(\phi_i) = \sum_t \log f_i(y_t, \phi_i)$ and let $\phi_{i,ML} = sup_{\phi_i} l_i(\phi_i)$. Since $T^{-1} l_i(\phi_i) \to E(\log f_i(y_{[1,t]}, \phi_i))$, by the uniform law of large numbers, $\phi_{i,ML}$ is consistent for $\phi_{i,0} = arg \max_{\phi_i} E \log f_i(y_{[1,t]}, \phi_i)$, where the expectations are taken with respect to F. If F is absolutely continuous with respect to f_i:

$$E \log f_i(y_{[1,t]}, \phi_i) - E \log F(y_{[1,t]}, \psi) = -\int F(y_{[1,t]}, \psi) \log \frac{F(y_{[1,t]}, \psi)}{f_i(y_{[1,t]}, \phi_i)} d\mu(y_{[1,t]}) = -KL_i(\phi_i) \quad (7)$$

Hence, $\phi_{i,0}$ is also the minimizer of KL_i, the Kullback-Liebler divergence between F and f_i.

Let $s_t^i(\phi_i) = \bigtriangledown_{\phi_i} \ln f_i(y_t, \phi_i)$ be the score of observation t and let $h_t^i(\phi_i) = \bigtriangledown_{\phi_i} s_t^i(\phi_i)$. If the maximum is in the interior of Φ, $\sum_t s_t^i(\phi_i) = 0$. First order expanding we have $0 \approx T^{-0.5} \sum_t s_t^i(\phi_{i,0}) + T^{0.5} V_1^{-1}(\phi_{i,ML} - \phi_{i,0})$ where $V_1 = -E(h_t^i(\phi_{i,0})) = \bigtriangledown_\phi^2 KL_i(\phi_i)_{\phi_i=\phi_{i,0}}$. By the central limit theorem for uncorrelated observations $T^{-0.5}(\phi_{i,ML} - \phi_{i,0}) \sim N(0, V_1 V_2 V_1')$, where $V_2 = E(s_t^i(\phi_i) s_t^i(\phi_i)')_{\phi_i=\phi_{i,0}}$, with the standard correction for V_2, when $y_{[1,t]}$ are correlated.

In typical applications $s_t^i(\phi_i)$ are computed with the Kalman filter and are function of martingale difference processes (the shocks of the model). Thus, $\sum_t s_t^i(\phi_i) = 0$ holds. Further regularity conditions are needed for the arguments to hold precisely (see, e.g. Mueller, 2013).

The composite likelihood geometrically averages different $f_i(y_t, \phi_i)$, each of which is misspecified. Thus, the composite model is, in general, misspecified with density $g(y_{1t}, \ldots, y_{Kt}, \theta, \eta_1, \ldots \eta_K) \equiv g(y_t, \phi) = \prod_i f_i(y_{it}, \phi_i)^{\omega_i}$. Repeating the argument of the previous paragraph, and under regularity conditions discussed in Xu and Reid (2011), when ω_i are fixed, ϕ_{CL}, the composite likelihood estimator, is consistent for $\phi_{0,CL}$, the minimizer of the KL divergence between the g and F. Furthermore, the scaled difference between ϕ_{CL} and $\phi_{CL,0}$ has an asymptotic normal distribution with zero mean and covariance matrix $V_{CL} = V_{CL,1}V_{CL,2}V'_{CL,1}$ where $V_{CL,2} = E(s_{CL,t}(\phi)s_{CL,t}(\phi)')$, $V_{CL,1} = -E[\bigtriangledown_\phi s_{CL,t}(\phi)]$ and $s_{CL,t}(\phi) = \bigtriangledown_\phi \ln g(y_t, \phi)$, all evaluated at $\phi = \phi_{CL}$. When the sub-models have different sample size, one needs to let $\min T_i \to \infty$.

When the weights are random, the asymptotic distribution of ϕ_i depends on $(\omega_1, \ldots, \omega_K)$. If the estimators of $\omega_1, \ldots \omega_K$ converge to a fixed KL pseudo value $\omega_{10}, \ldots, \omega_{K0}$, and no ω_{i0} is on the boundary of the parameter space, asymptotic normality still holds but the standard errors for ϕ_{CL} need to be adjusted for the randomness in ω_i. As long as the Godambe matrix is block diagonal in (ϕ, ω), one can ignore this extra uncertainty for inferential purposes.

Appendix B: Issues in quasi-posterior estimation

Drawing ω_i in MCMC algorithm There are various ways to draw candidate weights ω_i, $i = 1, ..., K$. If K is small, an independent Dirichlet proposal works well. If K is large, one could first logistically transform the weights and use a random walk proposal for the transformed weights. This approach has the disadvantage that the proposal is no longer a multivariate random walk (in particular, it is no longer symmetric). Furthermore, one needs to compute the Jacobian of the mapping, which may be tedious to code and may lead to numerical instabilities because of non-linearities.

Our preferred approach is to use a proposal density which directly operates on the weights. We call it 'random-walk Dirichlet', since the expected value of the proposal is the last accepted draw. Denote by ω^a the last accepted vector of weights, by ω^p a proposal draw, and by $\lambda > 0$ a scalar regulating the variance of the proposal. The proposal density is Dirichlet, denoted by $p_D(\omega^p|\omega^a, \lambda)$, with parameter $\lambda\omega^a$. Its mean is independent of λ and equal to ω^a. The variance of any element of ω^p is a decreasing function of λ. In an initial adaptive phase, where draws are discarded before computing posterior quantities, we adjust λ so as to achieve a reasonable acceptance probability (20-30%). This proposal density is not symmetric, and thus the acceptance probability needs to be properly modified.

paragraphAsymptotic properties of MCMC estimators Let χ_{CL} be the maximum composite likelihood estimator of $\chi = (\theta, \eta_1, \ldots, \eta_K, \omega_1, \ldots, \omega_K)$ and let χ_p be the mode of the prior $p(\chi)$. Let both χ_{CL} and χ_p be in the interior of the parameter space. Let $h(\chi_{CL}) = -\nabla^2_\chi \log CL(\chi_{CL}|y_t)$ and $h(\chi_p) = -\nabla^2_\chi \log p(\chi_p)$. Expanding quadratically the composite posterior $p_{CL}(\chi|y_t)$ we have

$$\begin{aligned}
&\propto \exp\{\log CL(\chi_{CL}|y_t) - 0.5(\chi-\chi_{CL})^T h(\chi_{CL})(\chi-\chi_{CL}) + \log p(\chi_p) - 0.5(\chi-\chi_p)^T h(\chi_p)(\chi-\chi_p)\} \\
&\approx N(\hat{\chi}, h(\chi_{CL}, \chi_p)^{-1}) \qquad (8)
\end{aligned}$$

where $\hat{\chi} = h(\chi_{CL}, \chi_p)^{-1}(h(\chi_{CL})\chi_{CL} + h(\chi_p)\chi_p)$ and $h(\chi_{CL}, \chi_p) = h(\chi_{CL}) + h(\chi_p)$.

Under regularity conditions, $p(\chi)$ vanishes as $T \to \infty$. Then, almost surely, the strong law of large number implies that

$$T^{-1}h(\chi_{CL}, \chi_p) \;\to\; -E(\nabla^2 \log CL(\hat{\chi}_0|y_t)) \equiv H(\hat{\chi}_0) \qquad (9)$$

$$\hat{\chi} \;=\; (T^{-1}h(\chi_{CL}, \chi_p))^{-1}(T^{-1}h(\chi_{CL})\chi_{CL} + T^{-1}h(\chi_p)\chi_p) \to \hat{\chi}_0 \qquad (10)$$

Thus as $T \to \infty$ $p_{CL}(\chi|y_t) \approx N(\hat{\chi}_0, T^{-1}H(\hat{\chi}_0)^{-1})$. Sufficient conditions that insure the above are, for example, in Deblasi and Walker (2013). Rubio and Villaverde (2004) provide conditions which are somewhat easier to verify in practice.

When χ_{CL} is not in the interior of the parameter space, for example, because $\omega_i \to 0$, for some i, η_i may become non-identifiable from the composite likelihood and the above may not hold. If we let $p(\eta_i) = p(\eta_i|y_{0t})$, where y_{0t} is a training sample of size $\bar{T}$, letting both T an $\bar{T}$ go to infinity, we will have that (9)-(10) hold for identified parameters while for those η_i for which $\omega_i \to 0$, $p_{CL}(\eta_i|y_t, y_{0t}) \approx N(\hat{\eta}_{i0}, \bar{T}^{-1}H(\hat{\eta}_{i0})^{-1})$, where $\hat{\eta}_{i0}$ is, e.g., ML estimator for η_i in the training sample.

When weak identification problems are present, care must be exercised since, the properties of ω_i may deviate from the standard ones stated in the text.

Appendix C: Tilting vs composite predictors

We look for a predictive density $p(z|y)$ solving:

$$\hat{p} = arg \min_p KL(p(z|y), f(z, \phi)) \qquad (11)$$

where z is any future sequence of y and ϕ a vector of parameters, subject to the constraint

$$E_p\{\log \frac{f(z|y_t, \phi)}{f(z, \phi)}\} = E_{Z|Y=y}\{\log \frac{f(z|y_t, \phi)}{f(z, \phi)}\} \;\; t = 1, \ldots, T \qquad (12)$$

and the normalization $E_p(1) = 1$, where E_p is the expectation with respect to the density $p(z|y)$, and $f(z, \phi)$ is any preliminary density of z, for example, its marginal. In words, we seek for the predictive density which is closest in the KL sense to any preliminary density $f(z,\phi)$ and reproduces the same conditional expectation as the true density $f(z|y,\phi)$ on functions $\log\frac{f(z|y_t,\psi)}{f(z,\phi)}$. Note that when $f(z)$ is disregarded, the problem becomes one of maximizing the entropy -$E_p[\log p(z|y)]$, subject to the constraints (12). The solution is $\hat{p}(z|y) = f(z,\phi)\exp\{\sum_t \xi_t \log\frac{f(z|y_t,\phi)}{f(z,\phi)}\} - \kappa(y_t,\phi,\xi)\}$ where $\kappa(y_t,\phi,\xi)$ is a normalizing constant, ξ_t are the Lagrange multipliers on the constraints (12). $\hat{p}(z|y)$ has an exponential tilting format: we tilt $f(z,\phi)$ in the directions spanned by $\log\frac{f(z|y_i,\phi)}{f(z,\phi)}$. If $\xi_t \geq 0, \sum_t \xi_t \leq 1$, then $\hat{p}(z|y)$ is the scaled version of the composite predictive density derived in section 4.5 with $\omega_t = \xi_t$, $t = 1,\ldots,T$ and $\omega_0 = 1 - \sum_t \xi_t$, where ω_0 is the weight on $f(z,\phi)$. Note that in this setup, ω_t satisfies the following (score) equation:

$$\frac{\partial E_{z|Y=y}\log f_p(Z|y,\phi,\omega_t)}{\partial \omega_t} = 0, \quad t = 1,\ldots,T \tag{13}$$

Thus, it can be chosen to maximize the conditional expected logarithmic score (13).

Appendix D: The models of section 5.1

1) Basic model with quadratic preferences, constant interest rate, exogenous permanent and transitory income process. Let $G = 1+g$ be the growth rate of permanent income. Let $\tilde{c}_t = \frac{c_t}{y_t^P}; \tilde{a}_t = \frac{a_t}{y_t^P}, y_t = y_t^T y_t^P$. The log linearized conditions are

$$\begin{aligned}
\hat{\tilde{c}}_t &= \hat{e}_{2t+1} + \hat{\tilde{c}}_{t+1} && (14)\\
\hat{\tilde{a}}_t &= \frac{1}{\bar{a}/G + \bar{y}^T - \bar{c}}(\bar{a}/G\hat{\tilde{a}}_{t-1} - \bar{a}/G\hat{e}_{2t} + \bar{y}\hat{y}_t^T - \bar{c}\hat{\tilde{c}}_t) && (15)\\
\hat{y}_t^P &= \hat{y}_{t-1}^P + \hat{e}_{2t} && (16)\\
\hat{y}_t^T &= \rho\hat{y}_{t-1}^T + \hat{e}_{1t} && (17)\\
\hat{c}_t &= \hat{\tilde{c}}_t + \hat{y}_t^P && (18)\\
\hat{a}_t &= \hat{\tilde{a}}_t + \hat{y}_t^P && (19)
\end{aligned}$$

where c_t is consumption, a_t are savings, y_t is income,ρ the persistence of transitory income, $(1+r)$ the gross real rate of interest $(1+r)\beta = 1$, $\sigma_i, i = 1,2$ the standard deviation of the transitory and permanent income, and variables with a bar indicate steady state quantities.

2) Model with exponential utility, constant interest rate, exogenous permanent and transitory income process. The instantaneous utility function is $u(c) = \frac{-1}{\theta}\exp(-\theta c_t)$, where $\theta > 0$ is the coefficient of risk aversion. The log linearized equations are:

$$
\begin{aligned}
-\hat{c}_t &= -\hat{c}_{t+1} + \frac{1}{\theta\bar{c}}(\hat{\sigma}_t + \hat{e}_{2t}) && (20)\\
\hat{\hat{a}}_t &= \frac{1}{\frac{\bar{a}}{G*\bar{\sigma}} + \bar{y}^T - \bar{\hat{c}}}\left(\frac{\bar{a}}{G\bar{\sigma}}\hat{\hat{a}}_{t-1} - \frac{\bar{a}}{G\bar{\sigma}}\hat{e}_{2t} - \frac{\bar{a}}{G}\hat{\sigma}_t + \bar{y}^T\hat{y}_t^T - \bar{c}\hat{\hat{c}}_t\right) && (21)\\
\hat{y}_t^P &= \hat{y}_{t-1}^P + \hat{\sigma}_t + \hat{e}_{2t} && (22)\\
\hat{y}_t^T &= \rho_1\hat{y}_{t-1}^T + \hat{\sigma}_t + \hat{e}_{1t} && (23)\\
\hat{\sigma}_t &= \rho_2\hat{\sigma}_{t-1} + \hat{e}_{3t} && (24)\\
\hat{c}_t &= \hat{\hat{c}}_t + \hat{y}_t^P && (25)\\
\hat{a}_t &= \hat{\hat{a}}_t + \hat{y}_t^P && (26)
\end{aligned}
$$

where σ_t is the standard deviation of the permanent and transitory income shock, and ρ_2 the persistence of the volatility process.

3) RBC model with separable CRRA preferences, labor supply decisions, capital accumulation, endogenous interest rate, permanent and transitory technology shocks. Letting α be the share of capital in production, γ the risk aversion coefficient, δ the capital depreciation rate, η the inverse of the Frish elasticity of labor supply, and assuming that $\log e_{2t}$ has zero mean, the log-linearized conditions are

$$
\begin{aligned}
\gamma\hat{\hat{c}}_t + \eta\hat{N}_t &= \hat{\hat{Y}}_t - \hat{N}_t && (27)\\
-\gamma\hat{\hat{c}}_t &= (1-\gamma)\hat{e}_{2t+1} - \gamma\hat{\hat{c}}_{t+1} + \frac{r}{1+r}\hat{r}_{t+1} && (28)\\
\hat{r}_t &= \frac{\alpha}{1+r}(\hat{\hat{Y}}_t - \hat{\hat{K}}_{t-1}) && (29)\\
\hat{\hat{Y}}_t &= \alpha(\hat{\hat{K}}_{t-1}) + (1-\alpha)(\hat{N}_t + \hat{\zeta}_t^T) && (30)\\
\hat{\hat{Y}}_t &= \frac{\bar{c}}{\bar{Y}}\hat{\hat{c}}_t + \frac{\bar{K}}{\bar{Y}}\hat{\hat{K}}_t + \frac{(1-\delta)}{G}\frac{\bar{K}}{\bar{Y}}\hat{\hat{K}}_{t-1} - \frac{(1-\delta)}{G}\frac{\bar{K}}{\bar{Y}}\hat{e}_{2t+1} && (31)\\
\hat{\zeta}_t^P &= G + \hat{\zeta}_{t-1}^P + \hat{e}_{2t} && (32)\\
\hat{\zeta}_t^T &= \rho\hat{\zeta}_{t-1}^T + \hat{e}_{1t} && (33)\\
\hat{c}_t &= \hat{\hat{c}}_t + \hat{y}_t^P && (34)\\
\hat{k}_t &= \hat{\hat{k}}_t + \hat{y}_t^P && (35)\\
\hat{y}_t &= \hat{\hat{y}}_t + \hat{y}_t^P && (36)
\end{aligned}
$$

where k_t is the capital stock and N_t is hours, ζ_t the technology disturbance and ρ the persistence of its transitory component.

4) Model with two types of agents optimizers and Rule of thumb (ROT) consumers, constant interest rate, permanent and transitory income components. Let $1-\omega$ be the share of ROT consumers. The log linearized conditions are:

$$\begin{aligned}
-\gamma\hat{\hat{c}}_{1t} &= (1-\gamma)\hat{e}_{2t+1} - \gamma\hat{\hat{c}}_{1t+1} && (37)\\
\hat{\hat{c}}_t^{ROT} &= \hat{y}_t^T && (38)\\
\hat{\hat{a}}_t &= \frac{1}{\bar{a}/G + \bar{y}^T - \bar{c}}(\bar{a}/G\hat{\hat{a}}_{t-1} - \bar{a}/G\hat{e}_{2t} + \bar{y}\hat{y}_t^T - \bar{c}_1\hat{\hat{c}}_{1t}) && (39)\\
\hat{y}_t^P &= G + \hat{y}_{t-1}^P + \hat{e}_{2t} && (40)\\
\hat{y}_t^T &= \rho\hat{y}_{t-1}^T + \hat{e}_{1t} && (41)
\end{aligned}$$

$$\begin{aligned}
\hat{c}_{1t} &= \hat{\hat{c}}_{1t} + \hat{y}_t^P && (42)\\
\hat{c}_t^{ROT} &= \hat{\hat{c}}_t^{ROT} + \hat{y}_T^P && (43)\\
\hat{c}_t &= \omega\hat{c}_{1t} + (1-\omega)\hat{c}_t^{ROT} && (44)\\
\hat{a}_t &= \hat{\hat{a}}_t + \hat{y}_t^P && (45)
\end{aligned}$$

where γ is the coefficient of relative risk aversion and the superscript ROT indicate the variables of the agents which do not save. We calibrate $\omega = 0.2, (1+r) = 1.01$.

5) Model with two types of optimizing agents, liquidity and non-liquidity constrained, constant interest rate, permanent and transitory income components. Utility depends on durable and non-durable consumption, relative price of non-durable is exogenous. The log-linear conditions are:

$$\begin{aligned}
\hat{\hat{c}}_{1t} - \hat{\hat{d}}_{1t} &= \\
\frac{1}{\zeta_1}(\hat{p}_t - \frac{1-\delta}{1+r}\hat{p}_{t+1}) & && (46)\\
\hat{\hat{a}}_{1t} - \hat{\hat{d}}_{1t} - \hat{p}_t &= 0 && (47)\\
\bar{p}\bar{d}_1\delta(\hat{p}_t + \hat{\hat{d}}_{1t}) + \bar{c}_1\hat{\hat{c}}_{1t} + \bar{a}_1\hat{\hat{a}}_{1t} &= \\
(1+r)\bar{a}_1\hat{\hat{a}}_{1t-1} + \hat{y}_t^T - [(1+r)\bar{a}_1 - (1-\delta)\bar{p}\bar{d}_1)]\hat{e}_{2t} & && (48)\\
\hat{\hat{c}}_{2t} - \hat{\hat{d}}_{2t} &= \\
\frac{1}{\zeta_2}(\hat{p}_t(1+\psi(\beta_2(1+r)-1)) - \beta_2(1-\delta)\hat{p}_{t+1} &+ \\
(\gamma-1)\beta_2[\psi(1+r)-(1-\delta)](\bar{c}_2\hat{\hat{c}}_{2t+1} - \bar{c}_2\hat{\hat{c}}_{2t} - \bar{d}_2\hat{\hat{d}}_{2t+1} + \bar{d}_2\hat{\hat{d}}_{2t+1})) & && (49)\\
\bar{p}\bar{d}_2\delta(\hat{p}_t + \hat{\hat{d}}_{2t}) + \bar{c}_2\hat{\hat{c}}_{2t} + \bar{a}_2\hat{\hat{a}}_{2t} &= \\
(1+r)\bar{a}_2\hat{\hat{a}}_{2t-1} + \hat{y}_t^T - [(1+r)\bar{a}_2 - (1-\delta)\bar{p}\bar{d}_2)]\hat{e}_{2t} & && (50)\\
\frac{\bar{a}_2}{B}\hat{\hat{a}}_{2t} + \frac{1-\bar{a}_2}{B}(\hat{p}_t + \hat{\hat{d}}_{2t}) &= 0 && (51)
\end{aligned}$$

These equations have six unknowns $(\hat{\hat{c}}_{1t}, \hat{\hat{c}}_{2t}, \hat{\hat{d}}_{1t}, \hat{\hat{d}}_{2t}, \hat{\hat{a}}_{1t}, \hat{\hat{a}}_{2t})$, given y_t^T, y_t^P, p_t. The remaining equations are:

$$
\begin{aligned}
(\gamma-1)(\bar{c}_2\hat{\hat{c}}_{2t} - \bar{d}_2\hat{\hat{d}}_{2t}) &+ \\
\frac{1}{\beta_2(1+r)-1}(\beta_2(1+r)(\gamma-1)(\bar{c}_2\hat{\hat{c}}_{2t+1} - \bar{c}_2\hat{\hat{c}}_{2t} - \bar{d}_2\hat{\hat{d}}_{2t+1} + \bar{d}_2\hat{\hat{d}}_{2t})) &= \hat{\mu}_t \quad (52) \\
\hat{\hat{c}}_{it} + \hat{y}_t^P &= \hat{c}_{it} \quad (53) \\
\hat{\hat{a}}_{it} + \hat{y}_t^P &= \hat{a}_{it} \quad (54) \\
\hat{\hat{d}}_{it} + \hat{y}_t^P &= \hat{d}_{it} \quad (55) \\
\hat{y}_{t-1}^P + \hat{e}_{2t} &= \hat{y}_t^P \quad (56) \\
\rho_1\hat{y}_{t-1}^T + \hat{e}_{1t} &= \hat{y}_t^T \quad (57) \\
\rho_2\hat{p}_{t-1} + \hat{e}_{3t} &= \hat{p}_t \quad (58) \\
\hat{y}_t^P + \hat{y}_t^T &= \hat{y}_t \quad (59) \\
\omega\hat{c}_{1t} + (1-\omega)\hat{c}_{2t} &= \hat{c}_t \quad (60) \\
\omega\hat{a}_{1t} + (1-\omega)\hat{a}_{2t} &= \hat{a}_t \quad (61) \\
\omega\hat{d}_{1t} + (1-\omega)\hat{d}_{2t} &= \hat{d}_t \quad (62)
\end{aligned}
$$

where i=1,2, $1-\omega$ is the share of liquidity constrained consumers and γ the Cobb-Douglass share of non-durable good d_{it} in the utility. $\zeta_1 = (1 - \frac{1-\delta}{1+r}), \zeta_2 = (1 - \beta_2((1-\delta) - \psi(1+r)) - 1)$, ψ is the share of durable financiable with assets, $\beta_2 > \beta_1$ and $\hat{\mu}$ is the Lagrange multiplier on the liquidity constraint (in percentage deviation from steady states). We set $\omega = 0.2, \psi = 0.95, B = 0.05, (1+r) = 1.01$.

6) Ad-hoc model: CRRA preferences, permanent and transitory income, habit in consumption (δ=habit parameter), an additional shock to the asset accumulation equation (M_t) and three measurement errors in the observation equations .

The log linear conditions are

$$-\frac{\gamma}{1-\delta/G}(\hat{\tilde{c}}_t - \delta/G\hat{\tilde{c}}_{t-1} + \delta/G\hat{e}_{2t}) = (1-\gamma)\hat{e}_{2t+1} - \frac{\gamma}{1-\delta/G}(\hat{\tilde{c}}_{t+1} - \delta/G\hat{\tilde{c}}_t + \delta/G\hat{e}_{2t+1}) \quad (63)$$

$$\bar{a}\hat{\tilde{a}}_t = (\frac{(1+r)\bar{a}}{G}(\hat{\tilde{a}}_{t-1} - \bar{a}/G\hat{e}_{2t}) + (1+r)(\bar{y}\hat{y}_t^T - \bar{c}\hat{\tilde{c}}_t) + e_{3t}) \quad (64)$$

$$\hat{y}_t^P = G + \hat{y}_{t-1}^P + \hat{e}_{2t} \quad (65)$$

$$\hat{y}_t^T = \rho\hat{y}_{t-1}^T + \hat{e}_{1t} \quad (66)$$

$$\hat{M}_t = \hat{y}_t^P + \hat{e}_{3t} \quad (67)$$

$$\hat{c}_t = \hat{\tilde{c}}_t + \hat{y}_t^P \quad (68)$$

$$\hat{a}_t = \hat{\tilde{a}}_t + \hat{y}_t^P \quad (69)$$

The measurement equations are:

$$\tilde{c}_t = \hat{c}_t + u_{1t} \quad (70)$$

$$\tilde{y}_t = \hat{y}_t + u_{2t} \quad (71)$$

$$\tilde{a}_t = \hat{a}_t + u_{3t} \quad (72)$$

whereu_{jt} are iid measurement errors with variance $\sigma_j^2, j = 1, 2, 3$.

Appendix E: The models of section 5.2

1) Herbst and Schorfheide (2015) model

$$y_t = E_t(y_{t+1}) - \frac{1}{\tau}(R_t - E_t(\pi_{t+1}) - E_t(z_{t+1})) + g_t - E_t(g_{t+1}) \quad (73)$$

$$\pi_t = \beta E_t(\pi_{t+1}) + \kappa(y_t - g_t) \quad (74)$$

$$R_t = \rho_R R_{t-1} + (1-\rho_R)(\phi_1\pi_t + \phi_2(y_t - g_t)) + \varepsilon_{R,t} \quad (75)$$

$$z_t = \rho_z z_{t-1} + \varepsilon_{z,t} \quad (76)$$

$$g_t = \rho_g g_{t-1} + \varepsilon_{g,t} \quad (77)$$

where y_t is output, π_t inflation, R_t the nominal rate, z_t a technology shock, g_t a demand shock and $e_{R,t}$ a monetary policy shock.

2) Justiniano, Primiceri and Tambalotti (2010) model

$$
\begin{aligned}
\hat{y}_t &= \frac{y+F}{y}\left[\alpha\hat{k}_t+(1-\alpha)\hat{L}_t\right] && (78)\\
\hat{\rho}_t &= \hat{w}_t+\hat{L}_t-\hat{k}_t && (79)\\
\hat{s}_t &= \alpha\hat{\rho}_t+(1-\alpha)\hat{w}_t && (80)\\
\hat{\pi}_t &= \gamma_f E_t\hat{\pi}_{t+1}+\gamma_b\hat{\pi}_{t-1}+\kappa\hat{s}_t+\kappa\hat{\lambda}_{p,t} && (81)\\
\hat{\lambda}_t &= \frac{h\beta e^{\gamma}}{(e^{\gamma}-h\beta)(e^{\gamma}-h)}E_t\hat{c}_{t+1}-\frac{e^{2\gamma}+h^2\beta}{(e^{\gamma}-h\beta)(e^{\gamma}-h)}\hat{c}_t+\frac{he^{\gamma}}{(e^{\gamma}-h\beta)(e^{\gamma}-h)}\hat{c}_{t-1} && (82)\\
&+ \frac{h\beta e^{\gamma}\rho_z-he^{\gamma}}{(e^{\gamma}-h\beta)(e^{\gamma}-h)}\hat{z}_t+\frac{e^{\gamma}-h\beta\rho_b}{e^{\gamma}-h\beta}\hat{b}_t && (83)\\
\hat{\lambda}_t &= \hat{R}_t+E_t\left(\hat{\lambda}_{t+1}-\hat{z}_{t+1}-\hat{\pi}_{t+1}\right) && (84)\\
\hat{\rho}_t &= \chi\hat{u}_t && (85)\\
\hat{\phi}_t &= (1-\delta)\beta e^{-\gamma}E_t\left(\hat{\phi}_{t+1}-\hat{z}_{t+1}\right)+\left(1-(1-\delta)\beta e^{-\gamma}\right)E_t\left[\hat{\lambda}_{t+1}-\hat{z}_{t+1}+\hat{\rho}_{t+1}\right] && (86)\\
\hat{\lambda}_t &= \hat{\phi}_t+\hat{u}_t-e^{2\gamma}S''(\hat{\iota}_t-\hat{\iota}_{t-1}+\hat{z}_t)+\beta e^{2\gamma}S''E_t\left[\hat{\iota}_{t+1}-\hat{\iota}_t+\hat{z}_{t+1}\right] && (87)\\
\hat{k}_t &= \hat{u}_t+\hat{\bar{k}}_{t-1}-\hat{z}_t && (88)\\
\hat{\bar{k}}_t &= (1-\delta)e^{-\gamma}\left(\hat{\bar{k}}_{t-1}-\hat{z}_t\right)+\left(1-(1-\delta)e^{-\gamma}\right)(\hat{u}_t+\hat{\iota}_t) && (89)\\
\hat{w}_t &= \frac{1}{1+\beta}\hat{w}_{t-1}+\frac{\beta}{1+\beta}E_t\hat{w}_{t+1}-\kappa_w\hat{g}_{w,t}+ && (90)\\
&+ \frac{\iota_w}{1+\beta}\hat{\pi}_{t-1}+\frac{1+\beta\iota_w}{1+\beta}\pi_t+\frac{\beta}{1+\beta}E_t\hat{\pi}_{t+1}+ && (91)\\
&+ \frac{\iota_w}{1+\beta}z_{t-1}-\frac{1+\beta\iota_w-\rho_z\beta}{1+\beta}z_t+\kappa_w\hat{\lambda}_{w,t} && (92)\\
\hat{g}_{w,t} &= \hat{w}_t-\left(\nu\hat{L}_t+\hat{b}_t-\hat{\lambda}_t\right) && (93)\\
\hat{R}_t &= \rho_R\hat{R}_{t-1}+(1-\rho_R)\left[\phi_\pi\hat{\pi}_t+\phi_X(\hat{x}_t-\hat{x}^*_t)\right]+\phi_{dX}\left[(\hat{x}_t-\hat{x}_{t-1})-\left(\hat{x}^*_t-\hat{x}^*_{t-1}\right)\right]+\hat{\eta}_{mp,t} && (94)\\
\hat{x}_t &= \hat{y}_t-\frac{\rho k}{y}\hat{u}_t && (95)\\
\frac{1}{g}\hat{y}_t &= \frac{1}{g}\hat{g}_t+\frac{c}{y}\hat{c}_t+\frac{i}{y}\hat{\iota}_t+\frac{\rho k}{y}\hat{u}_t && (96)
\end{aligned}
$$

Additional References

Deblasi, P. and S. Walker (2013) Bayesian asymptotics with misspecified models. *Statistica Sinica*, 23, 169-187.

Lv, J. and J. Liu (2014) Model selection principles in misspecified models. *Journal of the Royal Statistical Society*, 76, part 1, 141-167

Pakel, C., Shephard N. and K. Sheppard (2011). Nuisance parameters, composite likelihoods and a panel of GARCH models. *Statistica Sinica*, 21, 307-329.

Pauli, F., Racugno, W., and L. Ventura (2011). Bayesian composite marginal likelihoods. *Statistica Sinica*, 21, 149-164.

J. Fernandez Villaverde and J. Rubio Ramirez (2004). Comparing dynamic general equilibrium models to data: A Bayesian approach.*Journal of Econometrics*, 123, 128-157.

Xu, X. and N. Reid (2011) On the robustness of the maximum composite likelihood estimator. *Journal of Statistical Planning and Inference*, 141, 3047-3054.