APPENDIX TO MAN-BITES-DOG BUSINESS CYCLES

FOR ON-LINE PUBLICATION

KRISTOFFER P. NIMARK

This Appendix contains additional material for the paper Man-bites-dog Business Cycles. The next section derives the equilibrium expressions for the beauty contest model from Section 3 of the main paper. This is followed by a detailed description of how to solve the dynamic business cycle model of Section 4. The last section describes the Multiple-block Metropolis-Hastings algorithm used to estimate the business cycle model and contains some convergence diagnostics for the Markov chains. The Matlab codes used to estimate the model and plot the figures in the paper are available at the author's web site www.kris-nimark.net.

1. Solving the model of Morris and Shin (2002)

Here we describe the steps required to solve the model of Morris and Shin (2002) that were omitted from the main text.

1.1. **Set up.** The model of Morris and Shin (2002) consists of a utility function U_j for agent j

$$U_j = -(1-r)(a_j - x)^2 - r\left(L_j - \overline{L}\right) \tag{1.1}$$

where a_j is the action taken by agent j and where

$$L_j \equiv \int (a_i - a_j)^2 \, di \tag{1.2}$$

and

$$\overline{L} \equiv \int L_i di \tag{1.3}$$

The first order condition for agent j is given by

$$a_j = (1-r) E[x \mid \Omega_j] + rE[\overline{a} \mid \Omega_j] \tag{1.4}$$

where $\overline{a}$ is the cross-sectional average action defined as

$$\overline{a} \equiv \int a_i di \tag{1.5}$$

Date: June 4, 2013. *Address:* CREI, Universitat Pompeu Fabra and Barcelona GSE, Ramon Trias Fargas 25-27, Barcelona 08005.
e-mail: knimark@crei.cat *web*: www.kris-nimark.net.

1.2. **Equilibrium.** Instead of using the method of undetermined coefficients employed by Morris and Shin in their original paper, we will use a method that explicitly expresses the average action $\overline{a}$ as a function of higher order expectations of x. Of course, the resulting expressions describing the equilibrium are the same regardless of solution method. However, first expressing the solution in terms of higher order expectations is more convenient since we want to solve the model using two different information sets. It also helps to demonstrate how and why a man-bites-dog information structure affect the equilibrium average action.

Start by taking averages of (1.4) to get

$$\overline{a} = (1-r)\, x^{(1)} + r\overline{a}^{(1)} \tag{1.6}$$

where $\overline{a}^{(1)}$ is the average expectation of the average action $\overline{a}$, i.e.

$$\overline{a}^{(1)} \equiv \int E\left[\overline{a} \mid \Omega_j\right] dj \tag{1.7}$$

Substituting the terms in (1.6) into (1.7) we get

$$\overline{a}^{(1)} = (1-r)\, x^{(2)} + r\overline{a}^{(2)} \tag{1.8}$$

By taking the average expectation of (1.8) and so on, gives a general expression for the k order expectation of the average action

$$\overline{a}^{(k)} = (1-r)\, \theta^{(k+1)} + r\overline{a}^{(k+1)} \tag{1.9}$$

Now, repeated substitution of (1.9) into (1.6) and simplifying gives the convergent sum

$$\overline{a} = (1-r)\sum_{k=1}^{\infty} r^{k-1} x^{(k)} \tag{1.10}$$

which is expression (3.5) in the main text. The expression (1.10) describes the average action regardless of whether the signal y is available or not.

1.3. **Higher order expectations when $S = 0$.** From Section 2 of the main text we know that agent j's conditional (first order) expectations of x when $S = 0$ is given by

$$E\left(x \mid \Omega_j^0\right) = g_0 x_j \tag{1.11}$$

where

$$g_0 \equiv \frac{\sigma_\varepsilon^{-2}}{\sigma_\varepsilon^{-2} + \sigma^{-2}}$$

Taking averages across agents we get the average first order expectation

$$\begin{aligned} x^{(1)} &= \int E\left(x \mid \Omega_j^0\right) dj && (1.12)\\ &= g_0 x + g_0 \int \varepsilon_j dj && (1.13)\\ &= g_0 x && (1.14) \end{aligned}$$

since $\int \varepsilon_j dj = 0$. To find agent j's second order expectation, use that the model consistent expectation of $x^{(1)}$ equals the expectation of $g_0 x$

$$\begin{aligned} E\left(x^{(1)} \mid \Omega_j^0\right) &= E\left(g_0 x \mid \Omega_j^0\right) && (1.15)\\ &= g_0 E\left(x \mid \Omega_j^0\right) && (1.16)\\ &= g_0^2 x_j && (1.17)\end{aligned}$$

Again, averaging across agents we get an expression for the average second order expectation about x

$$x^{(2)} = g_0^2 x \tag{1.18}$$

repeating the same steps to get the average expectation about $x^{(2)}$, and so on, yields a general expression for the k order expectation when $S = 0$

$$x^{(k)} = g_0^k x \tag{1.19}$$

which is the expression (3.7) in the main text.

1.4. **Higher order expectations when $S = 1$.** When agents observe both x_j and y, agent j' conditional expectation of x is given by

$$\begin{aligned} E\left(x \mid \Omega_j^1\right) &= E(x \mid y) + E(x \mid [x_j - E(x \mid y)]) && (1.20)\\ &= g_y y + g_x (x_j - g_y y) && (1.21)\end{aligned}$$

where

$$g_y \equiv \frac{\sigma_\eta^{-2}}{\sigma_\eta^{-2} + \gamma^{-1}\sigma^{-2}}, \quad g_x \equiv \frac{\sigma_\varepsilon^{-2}}{\sigma_\varepsilon^{-2} + \sigma_\eta^{-2} + \gamma^{-1}\sigma^{-2}} \tag{1.22}$$

Taking averages

$$\begin{aligned} x^{(1)} &= \int E\left(x \mid \Omega_j^1\right) dj && (1.23)\\ &= g_y y + g_x (x - g_y y) && (1.24)\end{aligned}$$

Agent j's expectation of the average first order expectation is then given by

$$\begin{aligned} E\left(x^{(1)} \mid \Omega_j^1\right) &= E\left(g_y y + g_x (x - g_y y) \mid \Omega_j^1\right) && (1.25)\\ &= g_y y + g_x (g_y y + g_x (x_j - g_y y) - g_y y) && (1.26)\\ &= g_y y + g_x (g_x (x_j - g_y y)) && (1.27)\\ &= g_y y + g_x^2 (x_j - g_y y) && (1.28)\end{aligned}$$

$$x^{(2)} = g_y y + g_x^2 (x - g_y y) \tag{1.29}$$

Repeating the procedure gives a general expression for the k order expectation

$$x^{(k)} = g_y y + g_x^k (x - g_y y) \tag{1.30}$$

which is expression (3.8) in the main text.

1.5. **The average action as a function of x and y.** Plugging in the expression (1.19) for the higher order expectations when $S=0$ into the average action (1.10) gives

$$\begin{aligned}\overline{a} &= (1-r)\sum_{k=1}^{\infty} r^{k-1} g_0^k x &(1.31)\\ &= \frac{(1-r)\,g_0}{1-rg_0}x &(1.32)\end{aligned}$$

which corresponds to expression (3.11) in the main text. Plugging in the expression (1.30) for the higher order expectations when $S=1$ into the average action (1.10) gives

$$\begin{aligned}\overline{a} &= (1-r)\sum_{k=1}^{\infty} r^{k-1}\left(g_y y + g_x^k\,(x-g_y y)\right) &(1.33)\\ &= \frac{(1-r)\,g_x}{1-rg_x}\,(x-g_y y)+g_y y &(1.34)\\ &= \frac{(1-r)\,g_x}{1-rg_x}x+\left(1-\frac{(1-r)\,g_x}{1-rg_x}\right)g_y y &(1.35)\end{aligned}$$

which corresponds to expression (3.12) in the main text.

2. Solving the dynamic business cycle model

This Appendix describes how to find the equilibrium dynamics of the model using an iterative version of the method of undetermined coefficients. The algorithm is similar to that described in more detail in Nimark (2011) but adjusted to allow for time varying information structure. We start by defining some useful vectors and matrices and by making (informed) conjectures about the functional form of the solution. Using these definitions and conjectures we can then describe the two main steps in the solution algorithm. These steps are (i) Finding an expression for the endogenous variables as a function of the state taking the law of motion of the state as given. (ii) Find the law of motion of the state, taking the function mapping the state into endogenous outcomes as given. A cookbook style recipe describing an iterative algorithm to find a fixed point for this problem ends the section.

2.1. **Definitions and conjectures.** Here, we present the definition of the state and the conjectured forms for how the state evolves over time and how the endogenous variables inflation and output depend on the state.

2.1.1. *The law of motion of the state.* First, define the exogenous state vector $\mathbf{x}_{j,t}$ as

$$\mathbf{x}_{j,t} \equiv \begin{bmatrix} a_{j,t} \\ d_{j,t} \end{bmatrix} \tag{2.1}$$

and note that

$$\int \mathbf{x}_{j,t} dj = \mathbf{x}_t \tag{2.2}$$

In order to make optimal decisions, agents will need to form higher order expectations about the exogenous state vector $\mathbf{x}_t$ and agent j's hierarchy of expectations is defined as

$$X_{j,t} \equiv \begin{bmatrix} \mathbf{x}_{j,t} \\ E\left[X_t^{(\overline{k}-1)} \mid \Omega_{j,t}\right] \end{bmatrix} \tag{2.3}$$

where

$$X_t^{(\overline{k}-1)} \equiv \begin{bmatrix} \mathbf{x}_t \\ \mathbf{x}_t^{(1)} \\ \vdots \\ \mathbf{x}_t^{(\overline{k}-1)} \end{bmatrix} \tag{2.4}$$

and

$$\mathbf{x}_t^{(k+1)} \equiv \int E\left[\mathbf{x}_t^{(k)} \mid \Omega_{j,t}\right] dj \tag{2.5}$$

The aggregate state X_t is defined as the cross-sectional average of the expectations hierarchy (2.3)

$$X_t \equiv \int X_{j,t} dj \tag{2.6}$$

We will conjecture (and later verify) that the state X_t follows a VAR process

$$X_t = M(s^t)X_{t-1} + N(s^t)\mathbf{u}_t : \mathbf{u}_t \sim N(0, I) \tag{2.7}$$

Below we will show how common knowledge of rational expectations can be used to derive the law of motion (2.7) for X_t.

2.1.2. *The endogenous variables as function of the state.* We will also conjecture (and later verify) that output and inflation can be written as linear functions of the aggregate hierarchy of expectations X_t, the lagged interest rate r_{t-1} and the current aggregate shocks $\mathbf{u}_t$

$$\begin{bmatrix} y_t \\ \pi_t \end{bmatrix} = G(s^t)X_t + G_r r_{t-1} + G_u \mathbf{u}_t \tag{2.8}$$

It will be convenient to partition $G(s^t)$ into row vectors

$$\begin{bmatrix} y_t \\ \pi_t \end{bmatrix} = \begin{bmatrix} G_y(s^t) \\ G_\pi(s^t) \end{bmatrix} X_t + G_r r_{t-1} + G_u \mathbf{u}_t \tag{2.9}$$

Solving the model implies finding the matrices $G(s^t), G_r, G_u, M(s^t)$ and $N(s^t)$.

2.2. **Inflation and output as functions of the state.** For a given law of motion $M(s^t)$ and $N(s^t)$ we can find $G(s^t)$ by iterating on the (vector) Euler equation determined by the Euler equation for island j consumption

$$c_{j,t} = E\left[c_{j,t+1} \mid \Omega_t^j\right] - r_t + E\left[\overline{\pi}_{\mathcal{B}j,t+1} \mid \Omega_t^j\right] + d_{j,t} \tag{2.10}$$

and the island j Phillips curve

$$\begin{aligned} \pi_{j,t} &= \lambda(1+\varphi\delta)(\pi_t - \pi_{j,t}) + \lambda c_{j,t} + \lambda\varphi y_t - \lambda 1(+\varphi)a_{j,t} \\ &\quad + \beta E\left(\pi_{j,t+1} \mid \Omega_t^j\right) + \lambda\xi_{j,t}^1 + \lambda\varphi\xi_{j,t}^2 \end{aligned} \tag{2.11}$$

(where we use the definitions of $\overline{p}_{\mathcal{B}j,t}$ and $y_{j,t}$ from the main text). We can then write the consumption Euler equation and the Phillips curve in vector form as

$$\begin{bmatrix} c_{j,t} \\ \pi_{j,t} \end{bmatrix} = A \int E\left(\begin{bmatrix} c_{t+1} \\ \pi_{t+1} \end{bmatrix} \mid \Omega_{j,t}\right) + B(s^t)X_t^j + C(s^t)X_t + G_r r_{t-1} + G_u \mathbf{u}_t \tag{2.12}$$

where the matrices $A, B(s^t), C(s^t), G_r$ and G_u are given by

$$A = \begin{bmatrix} 1 & 1 \\ 0 & \beta \end{bmatrix} \tag{2.13}$$

$$B(s^t) = \begin{bmatrix} e_2 \\ -\lambda(1+\varphi)e_1 \end{bmatrix} + \begin{bmatrix} 0 \\ -\lambda(1+\varphi\delta)e_2 G_j(s^t) \end{bmatrix} + \begin{bmatrix} 0 \\ \lambda e_1 G_j(s^t) \end{bmatrix} \tag{2.14}$$

$$\begin{aligned} C(s^t) &= \begin{bmatrix} \phi_\pi G_\pi(s^t) + \phi_y G_y(s^t) \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \lambda(1+\varphi\delta)G_\pi(s^t) \end{bmatrix} \\ &\quad + \begin{bmatrix} 0 \\ \lambda\varphi G_y(s^t) \end{bmatrix} + \begin{bmatrix} 0 \\ -\lambda(1+\varphi\delta)G_\pi(s^t) \end{bmatrix} \end{aligned} \tag{2.15}$$

$$G_r = -\begin{bmatrix} \phi_r(1-\phi_r)^{-1} & (\lambda+\lambda\varphi)\,\phi_r(1-\beta\phi_r)^{-1} \end{bmatrix}' \tag{2.16}$$

$$G_u = -\begin{bmatrix} (1-\phi_r)^{-1} & (\lambda+\lambda\varphi)\,(1-\beta\phi_r)^{-1} \end{bmatrix}' \begin{bmatrix} \sigma_r & 0 \end{bmatrix} e_r' \tag{2.17}$$

The matrices G_r and G_u was computed by direct forward substitution of the effect of the lagged interest rate r_{t-1} and the monetary policy shock u_t^r on future consumption and inflation rates. The row vector e_r' picks out the element of $\mathbf{u}_t$ that corresponds to the monetary policy shock u_t^r in the Taylor rule.

2.2.1. *Aggregation.* Since we are not interested in deriving the dynamics of prices on an individual island, we can proceed by taking averages across islands and use the conjectured form (2.8) to get

$$\begin{aligned} G\left(s^t\right)X_t &= \left(B(s^t)+C(s^t)\right)X_t \\ &\quad +\omega AG\left(s_1^{t+1}\right)M\left(s_1^{t+1}\right)HX_t \\ &\quad +(1-\omega)AG\left(s_0^{t+1}\right)M\left(s_0^{t+1}\right)HX_t \end{aligned} \tag{2.18}$$

The expression (2.18) uses that expected output and inflation is given by

$$\begin{aligned} E\left(\left[\begin{array}{c} y_{t+1} \\ \pi_{t+1} \end{array}\right] \mid \Omega_{j,t}\right) &= \omega AG\left(s_1^{t+1}\right)M\left(s_1^{t+1}\right)HX_t \\ &\quad +(1-\omega)AG\left(s_0^{t+1}\right)M\left(s_0^{t+1}\right)HX_t \end{aligned} \tag{2.19}$$

where s_n^{t+1} denotes the history s^{t+1} with $s_{t+1}=n$. That is, the expectations of period $t+1$ inflation and output have to be weighted by the probability that there will be a man-bites-dog signal available in the next period. Equating coefficients then implies that $G\left(s^t\right)$ must satisfy

$$\begin{aligned} G\left(s^t\right) &= \left(B(s^t)+C(s^t)\right) \\ &\quad +\omega AG\left(s_1^{t+1}\right)M\left(s_1^{t+1}\right)H \\ &\quad +(1-\omega)AG\left(s_0^{t+1}\right)M\left(s_0^{t+1}\right)H \end{aligned} \tag{2.20}$$

For given matrices $A, B(s^t), C(s^t), H$ and $M(s^t)$ the matrices $G\left(s^t\right)$ (there is one matrix $G\left(s^t\right)$ for each history s^t) can be found by iterating on (2.20). This will be one important component in the iterative algorithm to solve the model described below.

2.3. **The law of motion of the state.** We now describe how to find the law of motion for the state. We have conjectured above that it will take the form

$$X_t = M(s^t)X_{t-1} + N(s^t)\mathbf{u}_t : \mathbf{u}_t \sim N(0,I) \tag{2.21}$$

and the partly, this law of motion is exogenous. That is, the first two row are given by the law of motion for the exogenous aggregate states a_t and d_t

$$\mathbf{x}_t = \rho\mathbf{x}_{t-1} + \nu_t\mathbf{u}_t$$

where

$$\rho = \left[\begin{array}{cc} \rho_a & 0 \\ 0 & \rho_d \end{array}\right] \tag{2.22}$$

and

$$\nu_t = \begin{bmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_d^2 \end{bmatrix} \text{ if } s_t = 0 \tag{2.23}$$

$$\nu_t = \begin{bmatrix} \gamma\sigma_a^2 & 0 \\ 0 & \sigma_d^2 \end{bmatrix} \text{ if } s_t = 1 \tag{2.24}$$

In order to find the remaining component of $M(s^t)$ and $N(s^t)$ we need to derive a law of motion for the average hierarchy of higher order expectations.

2.3.1. *The filtering problem of agent j.* Since X_t is made up of agents' higher order expectations about $\mathbf{x}_t$, the endogenous part of $M(s^t)$ and $N(s^t)$ depend on the how agents update their higher order expectations. For given $M(s^t)$ and $N(s^t)$ and for given $G(s^t)$ the filtering problem of the agents can be described by the state space system made up of the law of motion (2.21) and the measurement equation

$$\mathbf{z}_{j,t} = D(s^t)X_t + \begin{bmatrix} R_u(s_t) & R_j(s_t) \end{bmatrix} \begin{bmatrix} \mathbf{u}_t \\ \mathbf{u}_{j,t} \end{bmatrix} \tag{2.25}$$

Since the man-bites-dog signal is not always available, the matrices $D(s^t)$ and $R_u(s_t)$ and $R_j(s_t)$ agents' measurement equation vary over time and are given by

$$D(s^t) = \begin{bmatrix} e_1 \\ e_2 \\ G_\pi(s^t) \\ G_y(s^t) + G_\pi(s^t) \\ \phi_y G_y(s^t) + \phi_\pi G_\pi(s^t) \end{bmatrix}, R_u(s_t) = \begin{bmatrix} \mathbf{0}_{4\times 4} & \mathbf{0}_{4\times 1} \\ 0 & \sigma_r \end{bmatrix}, \tag{2.26}$$

$$R_j(s_t) = \begin{bmatrix} \sigma_{j,a} & 0 & 0 & 0 & 0 \\ 0 & \sigma_{j,d} & 0 & 0 & 0 \\ 0 & 0 & \sigma_{\xi 1} & 0 & 0 \\ 0 & 0 & 0 & \sigma_{\xi 2} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{2.27}$$

when $s_t = 0$ and

$$D(s^t) = \begin{bmatrix} e_1 \\ e_2 \\ G_\pi(s^t) \\ G_y(s^t) + G_\pi(s^t) \\ \phi_y G_y(s^t) + \phi_\pi G_\pi(s^t) \\ e_1 \end{bmatrix}, R_u(s_t) = \begin{bmatrix} \mathbf{0}_{4\times 4} & \mathbf{0}_{4\times 1} & \mathbf{0}_{4\times 1} \\ \mathbf{0}_{1\times 4} & \sigma_r & 0 \\ \mathbf{0}_{1\times 4} & 0 & \sigma_\eta \end{bmatrix}, \tag{2.28}$$

$$R_j(s_t) = \begin{bmatrix} \sigma_{j,a} & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{j,d} & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{\xi 1} & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{\xi 2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{2.29}$$

when $s_t = 1$. Since the system is conditionally linear gaussian, agent j's state estimate

$$X_{j,t|t} \equiv E\left[X_t \mid \Omega_{j,t}\right] \tag{2.30}$$

is optimally given by the Kalman update equation

$$X_{j,t|t} = M(s^t)X_{j,t-1|t-1} + K(s^t)\left[\mathbf{z}_{j,t} - D(s^t)M(s^t)X_{j,t-1|t-1}\right] \tag{2.31}$$

The Kalman gain $K(s^t)$ is given by the standard formula for systems in which the structural disturbances are correlated with the measurement errors

$$K(s^t) = \left[P(s^t)D'(s^t) + N'(s^t)R(s_t)\right]\left[D(s^t)P(s^t)D'(s^t) + R(s_t)R'(s_t)\right]^{-1} \tag{2.32}$$

The matrix $R(s_t)$ is the square root of the common and island specific measurement error covariance matrix

$$R(s_t) \equiv \left[\begin{array}{cc} R_u(s^t) & R_j(s^t) \end{array}\right] \tag{2.33}$$

and $P(s^t)$ is defined as

$$P(s^t) = E\left(X_t - X_{j,t|t-1,s^t}\right)\left(X_t - X_{j,t|t-1,s^t}\right)' \tag{2.34}$$

with

$$X_{j,t|t-1,s^t} \equiv E\left(X_t \mid z_j^{t-1}, s^t\right)$$

That is, the state estimation error covariance $P(s^t)$ is the covariance of the state estimation errors "prior" to observing $z_{j,t}$ but "posterior" to observing s_t. The state estimation errors are thus conditionally normally distributed with covariance $P(s^t)$ given by

$$\begin{aligned} P(s^t) &= M(s^t)\left[\begin{array}{c} P(s^{t-1}) - [P(s^{t-1})D'(s^{t-1}) + N'(s^{t-1})R(s_t)] \times \\ [D(s^{t-1})P(s^{t-1})D'(s^{t-1}) + R(s_{t-1})R'(s_{t-1})]^{-1} \times \\ [P(s^{t-1})D'(s^{t-1}) + N'(s^{t-1})R(s_{t-1})]' \end{array}\right] M'(s^t) \\ &\quad + N(s^t)N(s^t) \end{aligned} \tag{2.35}$$

(The term in brackets is the posterior state estimation error in period $t-1$.) The expressions for $K(s^t)$ and $P(s^t)$ are standard formulas for the Kalman filter recursions for state space systems with time-varying (but known) parameters (see for instance Anderson and Moore 1979).

2.3.2. *The average expectation hierarchy.* Substituting in the expression for $\mathbf{z}_t(j)$ into (2.31) gives

$$\begin{aligned} X_{j,t|t} &= \left[I - K(s^t)D(s^t)\right]M(s^t)X_{j,t-1|t-1} \\ &\quad + K(s^t)D\left(s^t\right)\mathbf{X}_t + K(s^t)R\left(s^t\right)\left[\begin{array}{c} \mathbf{u}_t \\ \mathbf{u}_{j,t} \end{array}\right] \end{aligned} \tag{2.36}$$

Taking averages across agents and using the law of motion for X_t to get

$$\begin{aligned} X_{t|t} &= \left[I - K(s^t)D(s^t)\right]M(s^t)X_{t-1|t-1} \\ &\quad + K(s^t)D\left(s^t\right)\left[M(s^t)X_{t-1} + N(s^t)\mathbf{u}_t\right] \\ &\quad + K(s^t)R_u\left(s_t\right)\mathbf{u}_t \end{aligned} \tag{2.37}$$

since $\int \mathbf{u}_t(j)dj = 0$. Amending the result to the exogenous law of motion for $\mathbf{x}_t$ results in the law of motion for the entire system

$$\begin{aligned} \begin{bmatrix} \mathbf{x}_t \\ X_{t|t} \end{bmatrix} &= \begin{bmatrix} \rho & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{t-1} \\ X_{t-1|t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ K(s^t)D(s^t)M(s^t) \end{bmatrix} \begin{bmatrix} \mathbf{x}_{t-1} \\ X_{t-1|t-1} \end{bmatrix} \\ &+ \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \left[I - K(s^t)D(s^t)\right] M(s^t) \end{bmatrix} \begin{bmatrix} \mathbf{x}_{t-1} \\ X_{t-1|t-1} \end{bmatrix} \\ &+ \begin{bmatrix} \nu(s_t) \\ K(s^t)D(s^t)N(s^t) \end{bmatrix} \mathbf{u}_t + K(s^t)R_u\left(s_t\right)\mathbf{u}_t \end{aligned} \tag{2.38}$$

For each s^t we thus need to find a fixed point of

$$M\left(s^t\right) = \begin{bmatrix} \rho & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ K(s^t)D(s^t)M(s^t) \end{bmatrix} + \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \left[I - K(s^t)D(s^t)\right] M(s^t) \end{bmatrix} \tag{2.39}$$

$$N\left(s^t\right) = \begin{bmatrix} \nu(s_t) \\ K(s^t)D(s^t)N(s^t) \end{bmatrix} + K(s^t)R_u\left(s_t\right) \tag{2.40}$$

where the last row and/or columns of the matrices have been cropped to make the matrices conformable (i.e. implementing the approximation that expectations of order $k > \overline{k}$ are redundant).

2.4. Algorithm for finding the solution.

(1) Start by making initial guesses for the $2^{\mathcal{T}}$ different versions of the matrices $G\left(s^t\right), M(s^t)$ and $N(s^t)$. A good initial guess is to set them such that the dynamics of the initial guess are equivalent to that of the full information solution.
(2) For given matrices $G\left(s^t\right), G_r, G_u, M(s^t)$ and $N(s^t)$ compute $2^{\mathcal{T}}$ "new" $G\left(s^t\right)$ using (2.20). (That is, one need to loop through the $2^{\mathcal{T}}$ different matrices $G\left(s^t\right)$.
(3) For given matrices $G_r, G_u, M(s^t)$ and $N(s^t)$ and the "new" $G\left(s^t\right)$ compute $2^{\mathcal{T}}$ "new" $M(s^t)$ and $N(s^t)$ using (2.39) and (2.40)
(4) Iterate on 2 and 3 until convergence.

To keep track of the $2^{\mathcal{T}}$ different versions of the matrices $G\left(s^t\right), M(s^t)$ and $N(s^t)$ it is helpful to use the following indexing strategy. Define $m \times n \times 2^{\mathcal{T}}$ arrays such that each "slice" of the array correspond to the matrices $G\left(s^t\right), M(s^t)$ and $N(s^t)$ for a given history s^t. Each unique finite history of s_t can be expressed as a decimal number using the mapping between binary and decimal numbers. For example, with $\mathcal{T} = 4$, the history

$$s^t = \{0, 0, 0, 1\} \tag{2.41}$$

will thus be assigned the index number 1 since the binary number 0001 equals 1 in the decimal system. Similarly, the history

$$s^t = \{1, 1, 0, 1\} \tag{2.42}$$

is given the index number 13 since the binary number 1101 equals 13 in the decimal system, and so on.[1]

[1]For instance, if only the last four periods matter for current dynamics we have $2^4 = 16$ different endogenous regimes, but if the last 8 periods matter the number of endogenous regimes is $2^8 = 256$. The endogenous regimes are quite similar which reduces the computational burden. Setting $\mathcal{T} = 12$ which implies that there

3. Estimating the model

The solved model can be represented in state space form as

$$Z_t = \overline{D}(s^t)X_t + D_r r_{t-1} + \overline{R}(s^t)\mathbf{u}_t \tag{3.1}$$
$$X_t = M(s^t)X_{t-1} + N(s^t)\mathbf{u}_t : \mathbf{u}_t \sim N(0, I) \tag{3.2}$$

where $\overline{D}(s^t)$ is a matrix that maps the state X_t into the vector Z_t that contains the time series used for estimation. The matrices $M(s^t), N(s^t), \overline{D}(s^t), D_r$ and $\overline{R}(s^t)$ are functions of the vector of parameters Θ. The posterior distribution of the model parameters Θ and the history of man-bites-dog indicators s^T can be estimated using the Multiple-block Metropolis-Hastings algorithm described in Chib (2001) by alternatingly sampling from the following two blocks of parameters.

3.1. Block 1: Sampling from $p\left(\Theta \mid Z^T, s^T\right)$.

By the method of composition we can express the conditional distribution $p\left(\Theta \mid Z^T, s^T\right)$ as

$$p\left(\Theta \mid Z^T, s^T\right) = \frac{p\left(Z^T \mid s^T, \Theta\right) p\left(\Theta \mid s^T\right)}{p\left(Z^T \mid s^T\right)} \tag{3.3}$$

implying the proportional relationship

$$p\left(\Theta \mid Z^T, s^T\right) \propto p\left(Z^T \mid s^T, \Theta\right) p\left(\Theta \mid s^T\right) \tag{3.4}$$

since $p\left(Z^T \mid s^T\right)$ does not depend on Θ. The first term on the right hand side of (3.4) is the likelihood function (defined below). The second term is proportional to $p\left(s^T \mid \Theta\right) p(\Theta)$, i.e.

$$p\left(\Theta \mid s^T\right) \propto p\left(s^T \mid \Theta\right) p(\Theta)$$

where $p\left(s^T \mid \Theta\right)$ is the pmf of $T + \mathcal{T} - 1$ repeated draws from independent Bernoulli trials with probability of success given by ω. That is,

$$\begin{aligned} p\left(s^T \mid \Theta\right) &= \prod_{t=-\mathcal{T}+1}^{T} p\left(s_t \mid \omega\right) \\ &= \omega^{\left(\sum_{t=-\mathcal{T}+1}^{T} s_t\right)} \times (1-\omega)^{\left((T+\mathcal{T}-1)-\sum_{t=-\mathcal{T}+1}^{T} s_t\right)} \end{aligned}$$

where $T + \mathcal{T} - 1$ is the dimension of s^T. The conditional target distribution $p\left(\Theta \mid Z^T, s^T\right)$ is thus proportional to the product of three distributions that are straightforward to evaluate for given Θ, Z^T and s^T, i.e.

$$p\left(\Theta \mid Z^T, s^T\right) \propto p\left(Z^T \mid s^T, \Theta\right) p\left(s^T \mid \Theta\right) p\left(\Theta\right) \tag{3.5}$$

We can then use the right hand side of the expression (3.5) to sample from $p\left(\Theta \mid Z^T, s^T\right)$ with a Metropolis step.

---

are $2^{12} = 4096$ different endogenous regimes result in a computational time of about 10 seconds to solve the model once on a standard desktop PC. This can be compared to a solution time of about 0.3 seconds for $\mathcal{T} = 5$ at the posterior mode of the estimated model.

3.2. **Block 2: Sampling from $p\left(s^T \mid Z^T, \Theta\right)$.** Similarly, to sample from $p\left(s^T \mid Z^T, \Theta\right)$ we will use that

$$p\left(s^T \mid Z^T, \Theta\right) = \frac{p\left(Z^T \mid s^T, \Theta\right) p\left(s^T \mid \Theta\right)}{p\left(Z^T \mid \Theta\right)} \tag{3.6}$$

implying the proportional relationship

$$p\left(s^T \mid Z^T, \Theta\right) \propto p\left(Z^T \mid s^T, \Theta\right) p\left(s^T \mid \Theta\right) \tag{3.7}$$

Again, $p\left(Z^T \mid s^T, \Theta\right) p\left(s^T \mid \Theta\right)$ is the product of the likelihood function and the pmf of $T + \mathcal{T} - 1$ repeated draws from independent Bernoulli trials and thus also straightforward to evaluate for given given $\Theta, Z^T$and s^T. The expression (3.7) can then be use to sample from $p\left(s^T \mid Z^T, \Theta\right)$ by a Metropolis step.

3.3. **A Multiple-Block Metropolis-Hastings Algorithm.** The Multiple-Block Metropolis-Hastings algorithm is implemented through the following steps.

(1) Specify initial values Θ_0 and s_0^T.
(2) Repeat for $j = 1, 2, ..., J$
  (a) Block 1: Draw Θ_j from $p\left(\Theta \mid s_{j-1}^T, Z^T\right)$
    (i) Generate the candidate parameter vector Θ^* from $q_\Theta\left(\Theta^* \mid \Theta_{j-1}\right)$
    (ii) Calculate $\alpha_j^\Theta = \min\left\{\frac{L\left(Z^T|s_{j-1}^T,\Theta^*\right)p\left(s^T|\Theta^*\right)p(\Theta^*)q_\Theta(\Theta_{j-1}|\Theta^*)}{L\left(Z^T|s_{j-1}^T,\Theta_{j-1}\right)p(s^T|\Theta_{j-1})p(\Theta_{j-1})q_\Theta(\Theta^*|\Theta_{j-1})}, 1\right\}$
    (iii) Set $\Theta_j = \Theta^*$ if $U(0,1) \leq \alpha_j^\Theta$ and $\Theta_j = \Theta_{j-1}$ otherwise.
  (b) Block 2: Draw s_j^T from $p\left(s^T \mid \Theta_j, Z^T\right)$
    (i) Generate s^{*T} from the proposal density $q_S\left(s^{*T} \mid s_{j-1}^T\right)$
    (ii) Calculate $\alpha_j^s = \min\left\{\frac{L\left(Z^T|s^{*T},\Theta_j\right)p\left(\Theta_j|s^{*T}\right)q_S\left(s_{j-1}^T|s^{*T}\right)}{L\left(Z^T|s_{j-1}^T,\Theta_j\right)p\left(\Theta^*|s_{j-1}^T\right)q_S\left(s^{*T}|s_{j-1}^T\right)}, 1\right\}$
    (iii) Set $s_j^T = s^{*T}$ if $U(0,1) \leq \alpha_j^s$ and $s_j^T = s_{j-1}^T$ otherwise.
(3) Return values $\{\Theta_0, \Theta_1,...,\Theta_J\}$ and $\{s_0^T, s_1^T,...,s_J^T\}$

In addition to the starting values Θ_0 and s_0^T we also need to choose a maximum order of expectation $\overline{k}$ and the maximum lag $\mathcal{T}$ for s_t. It is not computationally feasible to verify that the chosen values of these hyper parameters are large enough for each draw in the estimation algorithm. Therefore, $\overline{k}$ and $\mathcal{T}$ should be chosen to ensure some redundancy. This can be done by finding what the required $\overline{k}$ and $\mathcal{T}$ are for a "worst case" calibration of the model, i.e. finding a calibration that maximizes the importance of higher order expectations and lagged regimes and making sure that $\overline{k}$ and $\mathcal{T}$ are sufficiently large for those cases. The estimates reported in the paper are based on $\overline{k} = 8$ and $\mathcal{T} = 5$.

The prior distribution $p(\Theta)$ is described in the main text. The remaining components in the expressions in the algorithm are now described in detail.

3.4. **The likelihood function.** The log of the likelihood function $L\left(Z^T \mid s^T, \Theta\right)$ can be evaluated as

$$\log L\left(Z^T \mid s^T, \Theta\right) = -\frac{1}{2}\left\{\sum_{t=1}^{T} 2\pi \dim(Z_t) + \log\left|\Sigma(s^t)\right| + \widetilde{Z}_t'\Sigma(s^t)^{-1}\widetilde{Z}\right\} \tag{3.8}$$

where $\widetilde{Z}_t$ is the innovation to the observation vector Z_t and defined as

$$\widetilde{Z}_t \equiv Z_t - E\left(Z_t \mid Z^{t-1}\right) \tag{3.9}$$

The innovation covariance matrix $\Sigma(s^t)$ is defined as

$$\Sigma(s^t) \equiv E\left(\widetilde{Z}_t \widetilde{Z}_t'\right) \tag{3.10}$$

The innovation $\widetilde{Z}_t$ can for a given history s^t and parameter vector Θ be computed as the Kalman filter innovations to the state space system made up of the measurement equation (3.1) and the state equation (3.2).

There are four observable variables in Z_t that map into aggregate variables in the theoretical model, but in periods when there is no man-bites-dog signal available there are only three aggregate disturbances, i.e. the innovations to common productivity, the demand shocks and to the interest rate rule. To avoid stochastic singularity when evaluating the likelihood function, we treat the TFP series as well as the in CPI inflation and real GDP growth as being measured with (very small) noise.[2]

3.5. **The proposal density** $q_\Theta\left(\Theta^* \mid \Theta\right)$**.** In order to make the algorithm described above operational, we need to be specific about the proposal densities $q_\Theta\left(\Theta^* \mid \Theta_{j-1}\right)$ and $q_S\left(s^{*T} \mid s_{j-1}^T\right)$. Following Haario, Saksman, and Tamminen (2001), an Adaptive Random-Walk proposal density of the form

$$\Theta^* \sim N\left(\Theta_{j-1}, \Sigma_j\right) \tag{3.11}$$

was used to generate the candidate Θ^*. The covariance Σ_j is set to

$$\Sigma_j = c\frac{1}{j}\sum_{l=1}^{j} \Theta_l \Theta_l' \tag{3.12}$$

with the (constant) scalar c tuned to achieve an acceptance rate of approximately 22 per cent. (To initialize the adaptive component we set a constant Σ_j for with diagonal elements proportional to $|\Theta_0|$ for $j < 100$.) Since the proposal density is symmetric, i.e. since

$$q_\Theta\left(\Theta^* \mid \Theta_{j-1}\right) = q_\Theta\left(\Theta_{j-1} \mid \Theta^*\right) \tag{3.13}$$

the ratio

$$\frac{q_\Theta\left(\Theta_{j-1} \mid \Theta^*\right)}{q_\Theta\left(\Theta^* \mid \Theta_{j-1}\right)} = 1 \tag{3.14}$$

for all Θ_{j-1} and Θ^*. There is thus no need to compute ratio of proposal densities in Step 2.a.ii above.

[2]The variances of the measurement errors are about $1/100^{th}$ of the variance of the actual time series.

3.6. **The proposal density** $q_S\left(s^{*T} \mid s^T\right)$**.** The candidate s^{*T} is generated so that for each $t = 1, 2, 3, ..., T$ the probability that $s^*_t \neq s_{t,j-1}$ equals p^*. That is, if $s_{t,j-1} = 1$ the corresponding element in the candidate vector is set to 0 with probability p^* and kept unchanged at 1 with probability $(1 - p^*)$. If $s_{t,j-1} = 0$ the corresponding element in the candidate vector is set to 1 with probability p^* and kept unchanged at 0 with probability $(1 - p^*)$ .This proposal density implies that $q\left(s^{*T} \mid s^T\right)$ is a function only of the number of elements that differ between s^{*T} and s^T and is therefore symmetric, i.e. $q\left(s^{*T} \mid s^T\right) = q\left(s^T \mid s^{*T}\right)$. Again, there is thus no need to compute the ratio of the proposal densities in Step 2.b.ii above. The proposal density attached a positive probability to all possible realizations of s^T and thus satisfies a sufficient condition for convergence of the Markov chain to the target distribution (See Theorem 4.5.5 in Geweke 2005).

I experimented with alternative proposal densities for s^{*T}, including generating the candidate s^{*T} independently from the previous draw (i.e. updating the second block by an *independence M-H step*) and by updating s^T element by element as well as in a randomized order. In practise, the posterior estimate of s^T converges relatively rapidly compared to the posterior estimate of Θ regardless of the choice of specification of $q\left(s^{*T} \mid s^T\right)$. The exact choice of proposal density thus appears to have little impact on computational efficiency in terms of how many draws are needed to achieve convergence.

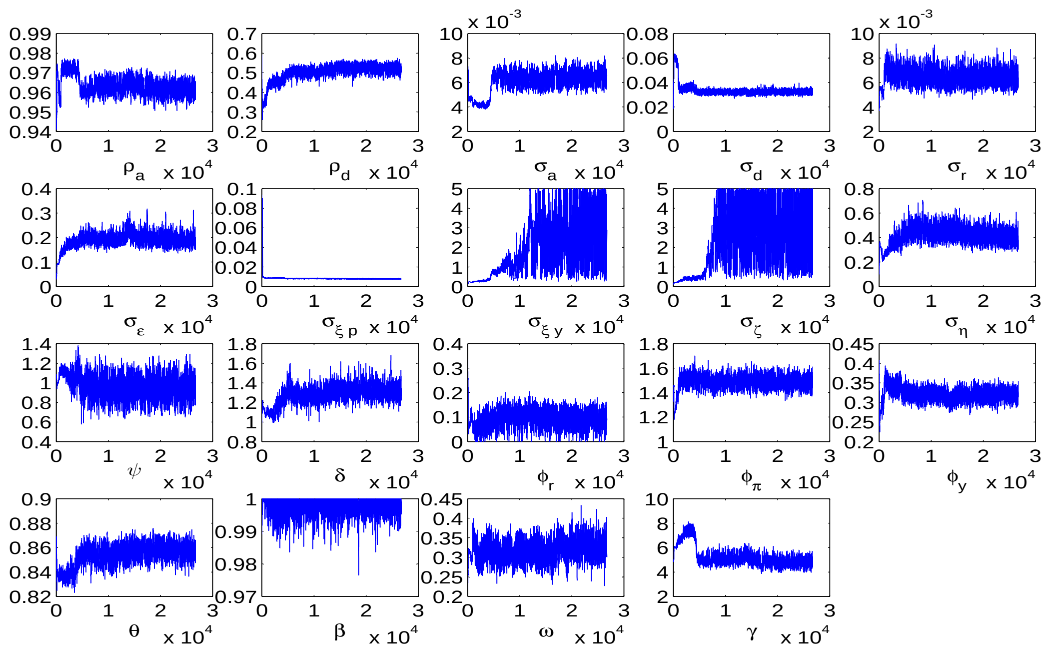

FIGURE 1. The raw Markov chain for Θ from the Metropolis algorithm.

3.7. **Convergence of the Markov Chain.** The Metropolis algorithm above was run for 2 600 000 draws and every 100th draw was saved. The time series of each element in the raw chain is plotted in Figure 1.

It is clear from inspecting Figure 1 that there were some large adjustments in the mean in the beginning of the chain. We therefore discard the first 1 000 000 draws of the Markov chain. The time series of the individual elements of the MCMC used for the figures and Table 1 in the paper are plotted in Figure 2 and Figure 3 plots the corresponding histograms for each element in Θ.

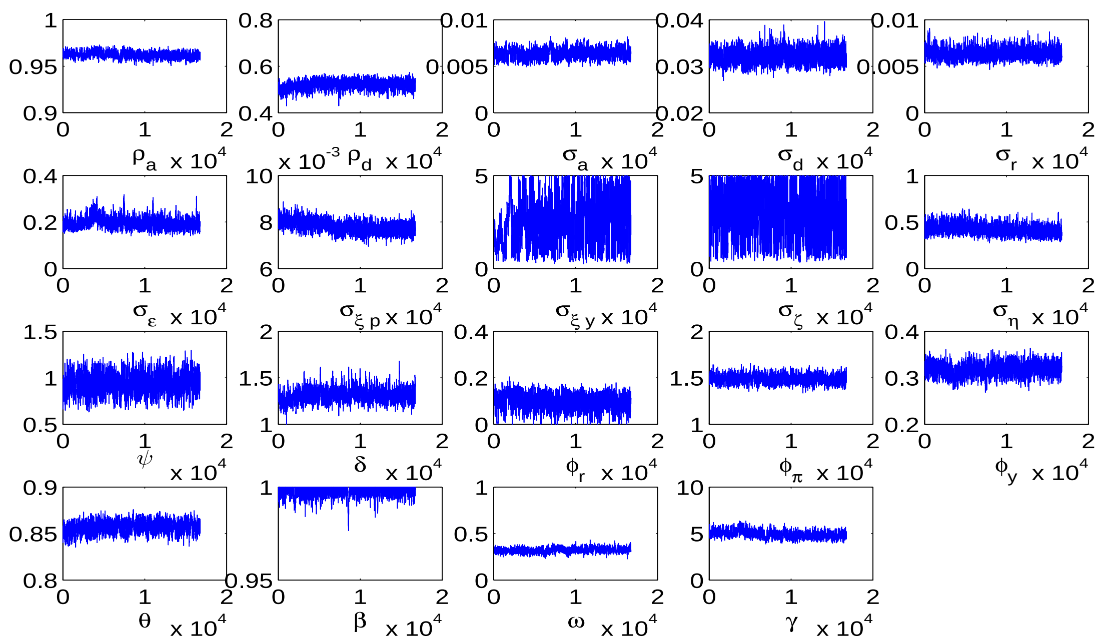

FIGURE 2. The segment of the Markov chain for Θ used for figures and tables in paper.

Most parameters appear well-identified, with the only two exceptions being $\sigma_{\xi y}$ and σ_{ζ}. The reason why it may be difficult to identify these two parameters using aggregate data is discussed in Section 5.8 of the main paper.

An informal but useful way to judge whether the MCMC has converged is to inspect the plots of the recursive means and variances of each element in the chain. To this end, Figure 4 contains the recursive mean of each element in the MCMC, i.e.

$$\frac{1}{j}\sum_{l=1}^{j}\Theta_l \tag{3.15}$$

for $j = 1, 2, ...$

Figure 5 contains the plots of the recursive sample variance of MCMC, i.e. the time series of elements from the diagonal of the matrix

$$\frac{1}{j}\sum_{l=1}^{j}\Theta_l\Theta_l' \tag{3.16}$$

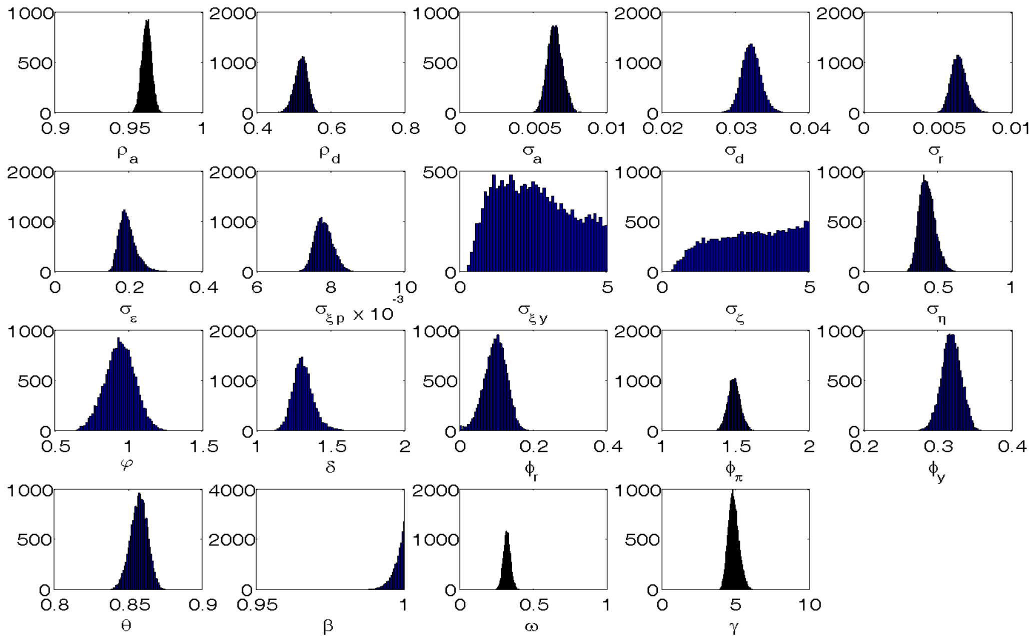

FIGURE 3. Histograms of elements in Markov chain for Θ.

for $j = 1, 2, ...$

In Figure 4 here are still some small adjustments in the mean of the chain even after 2 600 000 draws. However, plotting the recursive means for the last 500 000 draws in the chain as is done in Figure 6 suggest that also the means of the chain have indeed converged. (Using only the last 500 000 draws for the Figures in the paper does not have a material difference for any of the results.)

In general, the Markov chain for the posterior estimate of s^T converges much faster than the Markov chain for the posterior estimate of Θ. Figure 7 displays the mean of first and second half of the MCMC for s^T.

REFERENCES

[1] Anderson. B.D.O. and J.B. Moore, 1979, *Optimal Filtering*, reprinted by Dover Publications.

[2] Geweke, John, 2005, *Contemporary Bayesian Econometrics and Statistics*, Wiley Series in Probability and Statistics, John Wiley and Sons.

[3] Haario, H, E. Saksman and J. Tamminen, (2001), "An Adaptive Metropolis Algorithm", *Bernoulli*, 7(2), 223-242.

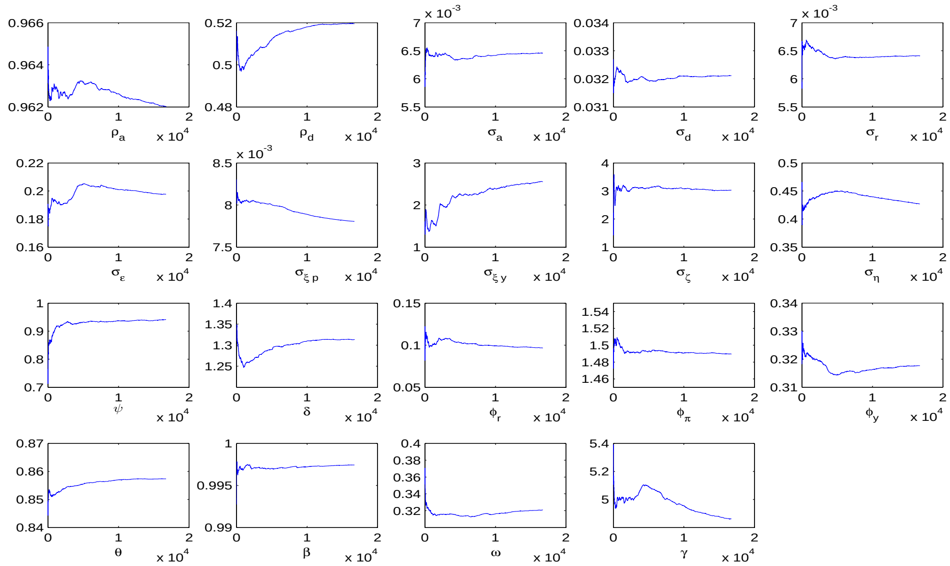

Figure 4. Recursive means of Markov chain for Θ.

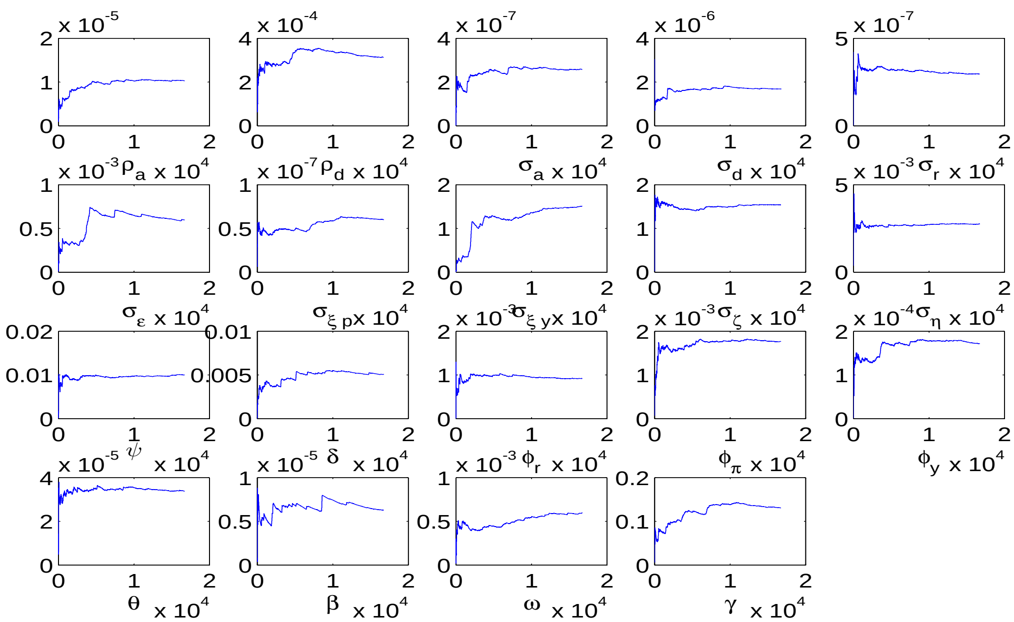

Figure 5. Recursive variances of Markov chain for Θ.

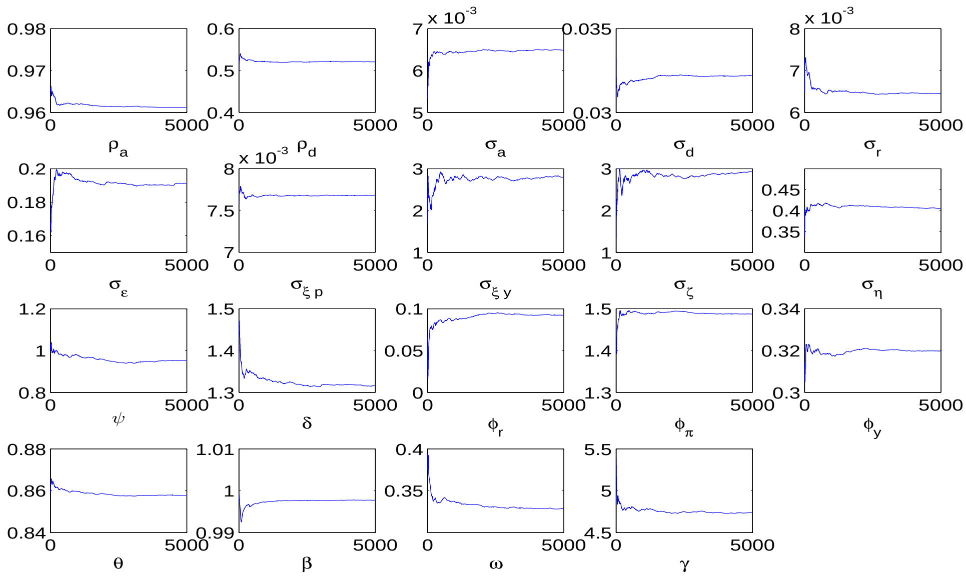

FIGURE 6. Recursive means of Markov chain for Θ for last 500 000 draws.

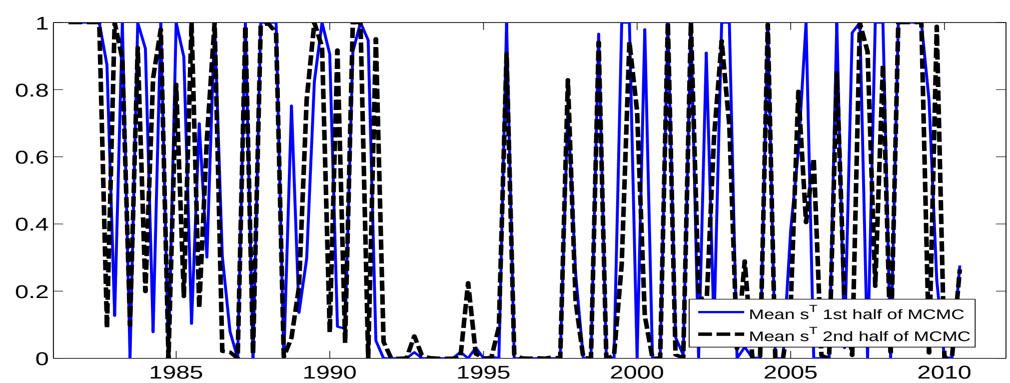

FIGURE 7. Mean of first and second half of Markov chain for s^T.