Parameter Identification in an Estimated New Keynesian Open Economy Model

Malin Adolfson
Sveriges Riksbank

Jesper Lindé*
Sveriges Riksbank and CEPR

This draft: November 9, 2007
First draft: October 18, 2006
Preliminary and incomplete, please do not circulate. Comments welcome

Abstract

In this paper, we use Monte Carlo methods to study the small sample properties of the classical maximum likelihood (ML) estimator in artificial samples generated by a New-Keynesian open economy DSGE model. The datagenerating process is the DSGE model estimated by Adolfson et al. (2006) using Bayesian techniques, and we employ an identical estimation strategy for the ML estimation on simulated data. Preliminary results suggest that the ML estimator works rather well and is unbiased for nearly all parameters and consistent for all parameters. However, the favorable properties of the ML estimator appear to be contingent on the inclusion of a sufficiently large set of observable variables. Finally, is also clear from the analysis that the data is more informative about some parameter than others, and that it harder to obtain convergence in the estimations if the initial guess in the optimizations is far from the true parameters.

Keywords: Identification; Maximum Likelihood estimation; New-Keynesian DSGE Model; Open economy.

JEL Classification Numbers: C13; C51; E30.

* *Address*: Sveriges Riksbank, SE-103 37 Stockholm, Sweden. *E-mails*: malin.adolfson@riksbank.se and jesper.linde@riksbank.se

The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Executive Board of Sveriges Riksbank.

1. Introduction

Following the seminal papers by Christiano, Eichenbaum and Evans (2005), and Smets and Wouters (2003), the interest to build and estimate dynamic stochastic general equilibrium (DSGE) models for welfare and policy analysis have increased sharply in both academic and policy surroundings. A recent paper by Canova and Sala (2006), however, suggests that it is difficult to ensure identification of parameters in DSGE models, casting doubts on the reliability of the results in the empirical DSGE literature. The models considered by Canova and Sala are stylized, and the number of estimated parameters is typically smaller than what has been estimated in the literature, e.g. in the seminal paper by Smets and Wouters (2003).

Most of the papers in the recent literature on estimated New-Keynesian type of DSGE models have used Bayesian estimation techniques. The choice of applying this approach can certainly be partly explained by compelling arguments why Bayesian methods are appropriate when thinking about macroeconomic models and policy applications. But there is also the possibility that Bayesian methods have been applied because "they work". If a given set of variables in the data set is not informative about some particular set of parameters in the model, i.e. if all parameters in the model are not identified by the data, the priors provide curvature for the posterior and thus enable "successful" estimation of the model.[1]

In this paper, we provide a study of the small sample properties of the classical maximum likelihood (ML) estimator in order to examine identification issues in the New-Keynesian small open economy DSGE model of Adolfson et al. (2006).[2] A log-linearized version of this DSGE model is used to generate artificial samples applying Adolfson et al's posterior median estimates of the parameters. The estimation strategy in the subsequent Monte Carlo exercise is essentially identical to the one adopted by Adolfson et al. (2006) with the exception that classical ML methods are used instead of Bayesian techniques. The key issue in the analysis is of course to understand whether identification is a generic problem for the new generation of DSGE models, or whether there are circumstances where DSGE models are identifed and can therefore (in principle) be successfully estimated with classical techniques. In addition, it is also well known that it is very difficult to conduct ML estimation of DSGE models on actual data. One possible explanation as to why classical ML methods appear to fail on actual data is that the DSGE models that we consider today are misspecified (see e.g. Del Negro et al., 2007 and Adolfson et al., 2006), and that some parameters are therefore driven to implausible values in reflection of the misspecification. By generating artifical samples where the DSGE model is in fact the data generating process, we know that there are no problems with misspecification, and we can then make an assessment of whether classical ML methods should be without problems. If this is the case, it would be tempting to draw the conclusion that the difficulties with ML estimation on actual data is due to problems with model misspecification rather that identification.

A limitation of our analysis is that it is restricted to one baseline model. So even if this

[1] A good hint about identification can be given by analyzing plots of the prior vs. the posterior; if the prior and posterior is identical for some parameters, this signals that those parameters are not properly identified. However, even if the prior equals the posterior, one cannot directly draw the conclusion that the parameter is not identified because it might be the case that the prior happens to coincide with what the data prefer. This latter possibility can of course be tested by changing the prior and redo the estimation, but that is not always done for all parameters in the recent empirical applications. However, even if the priors differ from the posteriors, it is not obvious that the model is identified. Suppose the following simple model $y_t = \frac{a_1}{a_2} y_{t-1} + e_t = \rho y_{t-1} + e_t$, where the econometrician buts two different priors on a_1 and a_2. From the data, ρ is identified but not a_1 and a_2 simultaneously, but if the priors are such that $a_1/a_2 \neq \rho$, the posteriors for both a_1 and a_2 will differ from their priors and both parameters will appear to be identifiable separately although they are actually not.

[2] With the exception for the uncovered interest rate parity condition, this model is essentially identical to the model originally developed by Adolfson et al. (2007).

particular model is identified, it does not allow us to draw general conclusions about identification in new Keynesian DSGE models. There are however three reasons why we think our analysis should be of interest nevertheless. First, we work with a model that has well-documented good empirical properties (see e.g. Adolfson et al. 2006). For the type of exercise that we conduct in the paper, we think this is of key importance as we can then make a case that our results are based on an empirically plausible model. Without imposing this restriction on the analysis, one could probably figure out examples of models that would lead to different conclusions than the one drawn here. The second reason why we think our analysis should be of particular interest, is that many models in the open economy literature are similar in spirit (see e.g. Cristadoro et al., 2007, Justiniano and Preston, 2006, Rabanal and Tuesta, 2005 and Smets and Wouters, 2002), and that many central banks are also currently working with similar models (e.g., FRB's SIGMA model (Erceg et al., 2006), ECB's NAWM model (Christoffel et al., 2007), and IMF's GEM model (Pesenti, 2003)). Third, given that the work of Canova and Sala (2006) focuses on the minimum distance estimator (by comparing impulse response functions in the model and a structural VAR), we add to their analysis by considering maximum likelihood estimation instead.

Preliminary results suggest that when an informative set of variables is included when matching the model to the data, most estimates are unbiased (with a few exceptions where the data is weakly informative). Moreover, when the sample size increases from 100 to 400 observations, the few cases where there are small sample biases disappear and the marginal distributions collapse around the true parameters. From this, we conjecture that all parameters are consistently estimated by classical ML techniques, and that all parameters in the DSGE model are identified given the set of variables we include in the estimation. But we also document in our analysis that some parameters suffer from problems with weak identfication.[3] The problem with weak identification pertains to some parameters in the policy rule, but more worrisome is that is a characteristic for the estimated degree of nominal wage stickiness in the model. [**To be continued.**]

Another interesting finding is that when we shrink the set of observed variables to include only "domestic" variables when estimating the model, we run into severe problems with weak identification of certain parameters as the parameter distributions pertaining specifically to the open economy aspects of the model become much more wide and also biased. This begs the need to consider which variables are required to be included among the observables if the purpose is to identify all parameters in the model.

The paper is organized as follows. In the next section, we desribe the open economy DSGE model that we use as the data generating process, and briefly describe how the model has been estimated on actual data. In Section 3, we describe how we estimate the model with classical ML techniques and how the small sample distribution of these estimates is obtained from the generated articfical data sets. In Section 4, we show the the results of the Monte Carlo exercises, with the aim to provide a better understanding of how to achieve improved identification of the model parameters, and why the classical ML estimator has poor properties for some parameters. Finally, we provide some concluding remarks in Section 5.

[3]Identification has to do with the ability to do inference about a particular set of model parameters given an observed set of variables. Following Canova and Sala (2006), we define a DSGE model to suffer from observational equivalence if different parameterizations of the model are indistinguishable from the point of view of the likelihood function. Another, perhaps more relevant case in practice, is a situationen there the DSGE model is plauged by weak identfication, i.e. where the likelihood function has a unique but weak curvature for (some of) the parameters that the econometrician tries to estimate. In the former case, the ML estimator will be inconsistent, wheras in the latter case, the ML estimator will be consistent but a very large sample is required to learn from aggregate data about (all) the parameters of the DSGE model.

2. The DGP - a New Keynesian open economy model

The model is an open economy DSGE model identical to the model presented and estimated in Adolfson et al. (2006). It shares its basic closed economy features with many recent new Keynesian models, including the benchmark models of Christiano, Eichenbaum and Evans (2005), Altig, Christiano, Eichenbaum and Lindé (2003), and Smets and Wouters (2003). This section gives an overview of the model and presents the key equations of it. We also discuss how the model is parameterized by reporting how it has been estimated on Swedish data by Adolfson et al. (2006) using Bayesian techniques.

2.1. The Model

The model economy includes four different categories of operating firms. These are domestic goods firms, importing consumption, importing investment, and exporting firms, respectively. Within each category there is a continuum of firms that each produces a differentiated good and set prices. The domestic goods firms produce their goods using capital and labour inputs, and sell them to a retailer which transforms the intermediate products into a homogenous final good that in turn is sold to the households. The final domestic good is a composite of a continuum of i differentiated goods, each supplied by a different firm, which follows the constant elasticity of substitution (CES) function

$$Y_t = \left[\int_0^1 (Y_{i,t})^{\frac{1}{\lambda_t^d}} \, di \right]^{\lambda_t^d}, 1 \leq \lambda_t^d < \infty, \tag{1}$$

where λ_t^d is a stochastic process that determines the time-varying flexible-price markup in the domestic goods market. The demand for firm i's differentiated product, $Y_{i,t}$, follows

$$Y_{i,t} = \left(\frac{P_{i,t}^d}{P_t^d} \right)^{-\frac{\lambda_t^d}{\lambda_t^d - 1}} Y_t. \tag{2}$$

The domestic production is exposed to unit root technology growth as in Altig et al. (2003). The production function for intermediate good i is given by

$$Y_{i,t} = z_t^{1-\alpha} \epsilon_t K_{i,t}^{\alpha} H_{i,t}^{1-\alpha} - z_t \phi, \tag{3}$$

where z_t is a unit-root technology shock capturing world productivity, ϵ_t is a domestic covariance stationary technology shock, $K_{i,t}$ the capital stock and $H_{i,t}$ denotes homogeneous labour hired by the i^{th} firm. A fixed cost $z_t \phi$ is included in the production function. We set this parameter so that profits are zero in steady state, following Christiano et al. (2005).

We allow for working capital by assuming that a fraction ν of the intermediate firms' wage bill has to be financed in advance through loans from a financial intermediary. Cost minimization yields the following nominal marginal cost for intermediate firm i:

$$MC_t^d = \frac{1}{(1-\alpha)^{1-\alpha}} \frac{1}{\alpha^{\alpha}} (R_t^k)^{\alpha} \left[W_t (1 + \nu (R_{t-1} - 1)) \right]^{1-\alpha} \frac{1}{(z_t)^{1-\alpha}} \frac{1}{\epsilon_t}, \tag{4}$$

where R_t^k is the gross nominal rental rate per unit of capital, R_{t-1} the gross nominal (economy wide) interest rate, and W_t the nominal wage rate per unit of aggregate, homogeneous, labour $H_{i,t}$.

Each of the domestic goods firms is subject to price stickiness through an indexation variant of the Calvo (1983) model. Since we have a time-varying inflation target in the model we allow for partial indexation to the current inflation target, but also to last period's inflation rate in order to allow for a lagged pricing term in the Phillips curve. Each intermediate firm faces in any period a probability $(1-\xi_d)$ that it can reoptimize its price. The reoptimized price is denoted $P_t^{d,new}$.[4] The different firms maximize profits taking into account that there might not be a chance to optimally change the price in the future. Firm i therefore faces the following optimization problem when setting its price

$$\max_{P_t^{d,new}} \mathrm{E}_t \sum_{s=0}^{\infty} (\beta\xi_d)^s \upsilon_{t+s}[((\pi_t^d \pi_{t+1}^d ... \pi_{t+s-1}^d)^{\kappa_d} (\bar{\pi}_{t+1}^c \bar{\pi}_{t+2}^c ... \bar{\pi}_{t+s}^c)^{1-\kappa_d} P_t^{d,new})Y_{i,t+s} \\ -MC_{i,t+s}^d(Y_{i,t+s} + z_{t+s}\phi^j)], \tag{5}$$

where the firm is using the stochastic household discount factor $(\beta\xi_d)^s \upsilon_{t+s}$ to make profits conditional upon utility. β is the discount factor, and υ_{t+s} the marginal utility of the households' nominal income in period $t+s$, which is exogenous to the intermediate firms. π_t^d denotes inflation in the domestic sector, $\bar{\pi}_t^c$ a time-varying inflation target of the central bank and $MC_{i,t}^d$ the nominal marginal cost.

The first order condition of the profit maximization problem in equation (5) yields the following log-linearized Phillips curve:

$$\begin{aligned}\left(\widehat{\pi}_t^d - \widehat{\bar{\pi}}_t^c\right) &= \frac{\beta}{1+\kappa_d\beta}\left(\mathrm{E}_t\widehat{\pi}_{t+1}^d - \rho_\pi\widehat{\bar{\pi}}_t^c\right) + \frac{\kappa_d}{1+\kappa_d\beta}\left(\widehat{\pi}_{t-1}^d - \widehat{\bar{\pi}}_t^c\right) \\ &\quad -\frac{\kappa_d\beta(1-\rho_\pi)}{1+\kappa_d\beta}\widehat{\bar{\pi}}_t^c + \frac{(1-\xi_d)(1-\beta\xi_d)}{\xi_d(1+\kappa_d\beta)}\left(\widehat{mc}_t^d + \widehat{\lambda}_t^d\right),\end{aligned} \tag{6}$$

where a hat denotes log-deviation from steady state (i.e., $\hat{X}_t = \ln X_t - \ln X$).

We now turn to the import and export sectors. There is a continuum of importing consumption and investment firms that each buys a homogenous good at price P_t^* in the world market, and converts it into a differentiated good through a brand naming technology. The exporting firms buy the (homogenous) domestic final good at price P_t^d and turn this into a differentiated export good through the same type of brand naming. The nominal marginal cost of the importing and exporting firms are thus $S_tP_t^*$ and P_t^d/S_t, respectively, where S_t is the nominal exchange rate (domestic currency per unit of foreign currency). The differentiated import and export goods are subsequently aggregated by an import consumption, import investment and export packer, respectively, so that the final import consumption, import investment, and export good is each a CES composite according to the following:

$$C_t^m = \left[\int_0^1 (C_{i,t}^m)^{\frac{1}{\lambda_t^{mc}}} di\right]^{\lambda_t^{mc}}, \qquad I_t^m = \left[\int_0^1 (I_{i,t}^m)^{\frac{1}{\lambda_t^{mi}}} di\right]^{\lambda_t^{mi}}, \qquad X_t = \left[\int_0^1 (X_{i,t})^{\frac{1}{\lambda_t^{x}}} di\right]^{\lambda_t^{x}}, \tag{7}$$

where $1 \le \lambda_t^j < \infty$ for $j = \{mc, mi, x\}$ is the time-varying flexible-price markup in the import consumption (mc), import investment (mi) and export (x) sector. By assumption the continuum of consumption and investment importers invoice in the domestic currency and exporters in the foreign currency. In order to allow for short-run incomplete exchange rate pass-through to import as well as export prices we therefore introduce nominal rigidities in the local currency

[4] For the firms that are not allowed to reoptimize their price, we adopt the indexation scheme $P_{t+1}^d = (\pi_t^d)^{\kappa_d} (\bar{\pi}_{t+1}^c)^{1-\kappa_d} P_t^d$ where κ_d is an indexation parameter.

price, following for example Smets and Wouters (2002). This is modeled through the same type of Calvo setup as above. The price setting problems of the importing and exporting firms are completely analogous to that of the domestic firms in equation (5), and the demand for the differentiated import and export goods follow similar expressions as to equation (2). In total there are thus four specific Phillips curve relations determining inflation in the domestic, import consumption, import investment and export sectors.

In the model economy there is also a continuum of households which attain utility from consumption, leisure and real cash balances. The preferences of household j are given by

$$\mathrm{E}_0^j \sum_{t=0}^{\infty} \beta^t \left[\zeta_t^c \ln\left(C_{j,t} - bC_{j,t-1}\right) - \zeta_t^h A_L \frac{(h_{j,t})^{1+\sigma_L}}{1+\sigma_L} + A_q \frac{\left(\frac{Q_{j,t}}{z_t P_t^d}\right)^{1-\sigma_q}}{1-\sigma_q} \right], \tag{8}$$

where $C_{j,t}$, $h_{j,t}$ and $Q_{j,t}/P_t^d$ denote the j^{th} household's levels of aggregate consumption, labour supply and real cash holdings, respectively. Consumption is subject to habit formation through $bC_{j,t-1}$, such that the household's marginal utility of consumption is increasing in the quantity of goods consumed last period. ζ_t^c and ζ_t^h are persistent preference shocks to consumption and labour supply, respectively. To make cash balances in equation (8) stationary when the economy is growing they are scaled by the unit root technology shock z_t. Households consume a basket of domestically produced goods and imported products which are supplied by the domestic and importing consumption firms, respectively. Aggregate consumption is assumed to be given by the following constant elasticity of substitution (CES) function:

$$C_t = \left[(1-\omega_c)^{1/\eta_c} \left(C_t^d\right)^{(\eta_c-1)/\eta_c} + \omega_c^{1/\eta_c} \left(C_t^m\right)^{(\eta_c-1)/\eta_c} \right]^{\eta_c/(\eta_c-1)}, \tag{9}$$

where C_t^d and C_t^m are consumption of the domestic and imported good, respectively. ω_c is the share of imports in consumption, and η_c is the elasticity of substitution across consumption goods.

The households invest in a basket of domestic and imported investment goods to form the capital stock, and decide how much capital to rent to the domestic firms given costs of adjusting the investment rate. The households can increase their capital stock by investing in additional physical capital (I_t), taking one period to come in action. The capital accumulation equation is given by

$$K_{t+1} = (1-\delta)K_t + \Upsilon_t \left(1 - \tilde{S}\left(I_t/I_{t-1}\right)\right) I_t, \tag{10}$$

where $\tilde{S}\left(I_t/I_{t-1}\right)$ determines the investment adjustment costs through the estimated parameter $\tilde{S}''$, and Υ_t is a stationary investment-specific technology shock. Total investment is assumed to be given by a CES aggregate of domestic and imported investment goods (I_t^d and I_t^m, respectively) according to

$$I_t = \left[(1-\omega_i)^{1/\eta_i} \left(I_t^d\right)^{(\eta_i-1)/\eta_i} + \omega_i^{1/\eta_i} \left(I_t^m\right)^{(\eta_i-1)/\eta_i} \right]^{\eta_i/(\eta_i-1)}, \tag{11}$$

where ω_i is the share of imports in investment, and η_i is the elasticity of substitution across investment goods.

Further, along the lines of Erceg, Henderson and Levin (2000), each household is a monopoly supplier of a differentiated labour service which implies that they can set their own wage. After having set their wage, households supply the firms' demand for labour at the going wage rate.

Each household sells its labour to a firm which transforms household labour into a homogenous good that is demanded by each of the domestic goods producing firms. Wage stickiness is introduced through the Calvo (1983) setup, with partial indexation to last period's CPI inflation rate, the current inflation target and the technology growth. Household j reoptimizes its nominal wage rate $W_{j,t}^{new}$ according to the following

$$\begin{array}{c}\max\limits_{W_{j,t}^{new}} \quad \mathrm{E}_t \sum_{s=0}^{\infty} (\beta \xi_w)^s \left[-\zeta_{t+s}^h A_L \frac{(h_{j,t+s})^{1+\sigma_L}}{1+\sigma_L} + \right. \\ \left. \upsilon_{t+s} \frac{(1-\tau_{t+s}^y)}{(1+\tau_{t+s}^w)} \left(\left(\pi_t^c ... \pi_{t+s-1}^c \right)^{\kappa_w} \left(\bar{\pi}_{t+1}^c ... \bar{\pi}_{t+s}^c \right)^{(1-\kappa_w)} \left(\mu_{z,t+1} ... \mu_{z,t+s} \right) W_{j,t}^{new} \right) h_{j,t+s} \right],\end{array} \tag{12}$$

where ξ_w is the probability that a household is not allowed to reoptimize its wage, τ_t^y a labour income tax, τ_t^w a pay-roll tax (paid for simplicity by the households), and $\mu_{z,t} = z_t / z_{t-1}$ is the growth rate of the permanent technology level.[5]

The households can accumulate capital, save in domestic and foreign bonds, and also hold cash. The choice between domestic and foreign bond holdings balances into an arbitrage condition pinning down expected exchange rate changes (i.e., an uncovered interest rate parity condition). To ensure a well-defined steady-state in the model, we assume that there is a premium on the foreign bond holdings which depends on the aggregate net foreign asset position of the domestic households, following, e.g., Lundvik (1992), and Schmitt-Grohé and Uribe (2001). Our specification of the risk premium also includes the expected change in the exchange rate $E_t S_{t+1}/S_{t-1}$ which is based on the vast empirical evidence of a forward premium puzzle in the data (i.e., that risk premia are strongly negatively correlated with the expected depreciation of the exchange rate), see e.g. Fama (1984) Duarte and Stockman (2005), an observation which is not consistent with a standard UIP condition. Our modification enables the model to induce endogenous persistence in the exchange rate and generates a hump-shaped response of the real exchange rate after a shock to monetary policy, see Adolfson et al. (2006) for a more detailed discussion. The risk premium is given by:

$$\Phi(a_t, S_t, \tilde{\phi}_t) = \exp\left(-\tilde{\phi}_a (a_t - \bar{a}) - \tilde{\phi}_s \left(\frac{E_t S_{t+1}}{S_t} \frac{S_t}{S_{t-1}} - 1 \right) + \tilde{\phi}_t \right), \tag{13}$$

where $a_t \equiv (S_t B_t^*)/(P_t z_t)$ is the net foreign asset position, and $\tilde{\phi}_t$ is a shock to the risk premium. The UIP condition in its log-linearized form is given by:

$$\widehat{R}_t - \widehat{R}_t^* = \left(1 - \widetilde{\phi}_s \right) E_t \Delta \widehat{S}_{t+1} - \widetilde{\phi}_s \Delta \widehat{S}_t - \widetilde{\phi}_a \widehat{a}_t + \widehat{\widetilde{\phi}}_t. \tag{14}$$

By setting $\tilde{\phi}_s = 0$ we obtain the UIP condition typically used in small open economy models (see, e.g., Adolfson et al., 2005a).

Following Smets and Wouters (2003), monetary policy is approximated with a generalized Taylor (1993) rule. The central bank is assumed to adjust the short term interest rate in response to deviations of CPI inflation from the time-varying inflation target, the output gap (measured as actual minus trend output), the real exchange rate $\left(\hat{x}_t \equiv \hat{S}_t + \hat{P}_t^* - \hat{P}_t^c \right)$ and the interest rate set in the previous period. The instrument rule (expressed in log-linearized terms) follows:

$$\begin{aligned}\widehat{R}_t &= \rho_{R,t} \widehat{R}_{t-1} + \left(1 - \rho_{R,t} \right) \left[\widehat{\bar{\pi}}_t^c + r_{\pi,t} \left(\hat{\pi}_{t-1}^c - \widehat{\bar{\pi}}_t^c \right) + r_{y,t} \hat{y}_{t-1} + r_{x,t} \hat{x}_{t-1} \right] \\ &\quad + r_{\Delta\pi,t} \Delta \hat{\pi}_t^c + r_{\Delta y,t} \Delta \hat{y}_t + \varepsilon_{R,t},\end{aligned} \tag{15}$$

[5] For the households that are not allowed to reoptimize, the indexation scheme is $W_{j,t+1} = (\pi_t^c)^{\kappa_w} \left(\bar{\pi}_{t+1}^c \right)^{(1-\kappa_w)} \mu_{z,t+1} W_{j,t}^{new}$, where κ_w is an indexation parameter.

where $\varepsilon_{R,t}$ is an uncorrelated monetary policy shock.

The structural shock processes in the model is given in log-linearized form by the univariate representation

$$\hat{\varsigma}_t = \rho_\varsigma \hat{\varsigma}_{t-1} + \varepsilon_{\varsigma,t}, \quad \varepsilon_{\varsigma,t} \overset{iid}{\sim} N\left(0, \sigma_\varsigma^2\right)$$

where $\varsigma_t = \{\ \mu_{z,t},\ \epsilon_t, \lambda_t^j,\ \zeta_t^c,\ \zeta_t^h,\ \Upsilon_t,\ \tilde{\phi}_t,\ \varepsilon_{R,t}, \bar{\pi}_t^c,\ \tilde{z}_t^*\}$ and $j = \{d, mc, mi, x\}$.

The government spends resources on consuming part of the domestic good, and collects taxes from the households. The resulting fiscal surplus/deficit plus the seigniorage are assumed to be transferred back to the households in a lump sum fashion. Consequently, there is no government debt. The fiscal policy variables - taxes on capital income, labour income, consumption, and the pay-roll, together with (HP-detrended) government expenditures - are assumed to follow an identified VAR model with two lags.

To simplify the analysis we adopt the assumption that the foreign prices, output (HP-detrended) and interest rate are exogenously given by an identified VAR model with four lags. Both the foreign and the fiscal VAR models are being estimated, using uninformative priors, ahead of estimating the structural parameters in the DSGE model.[6]

To clear the final goods market, the foreign bond market, and the loan market for working capital, the following three constraints must hold in equilibrium:

$$C_t^d + I_t^d + G_t + C_t^x + I_t^x \leq z_t^{1-\alpha} \epsilon_t K_t^\alpha H_t^{1-\alpha} - z_t \phi, \tag{16}$$

$$S_t B_{t+1}^* = S_t P_t^x \left(C_t^x + I_t^x\right) - S_t P_t^* \left(C_t^m + I_t^m\right) + R_{t-1}^* \Phi(a_{t-1}, \widetilde{\phi}_{t-1}) S_t B_t^*, \tag{17}$$

$$\nu W_t H_t = \mu_t M_t - Q_t, \tag{18}$$

where G_t is government expenditures, C_t^x and I_t^x are the foreign demand for export goods, and $\mu_t = M_{t+1}/M_t$ is the monetary injection by the central bank. When defining the demand for export goods, we introduce a stationary asymmetric (or foreign) technology shock $\tilde{z}_t^* = z_t^*/z_t$, where z_t^* is the permanent technology level abroad, to allow for temporary differences in permanent technological progress domestically and abroad.

To compute the equilibrium decision rules, we proceed as follows. First, we stationarize all quantities determined in period t by scaling with the unit root technology shock z_t. Then, we log-linearize the model around the constant steady state and calculate a numerical (reduced form) solution with the AIM algorithm developed by Anderson and Moore (1985).

2.2. Parameterization of the model on actual data

We start the empirical analysis by estimating the DSGE model on actual data, using a Bayesian approach and placing a prior distribution on the structural parameters. We use quarterly Swedish data for the period $1980Q1 - 2004Q4$. All data were taken from Statistics Sweden, except the repo rate which were taken from Sveriges Riksbank. The nominal wage is taken from Statistics Sweden and is deflated by the GDP deflator. The foreign variables on output, the

[6] The reason why we include foreign output HP-detrended and not in growth rates in the VAR is that the level of foreign output enters the DSGE model (e.g., in the aggregate resource constraint). In the state-space representation of the model, which links the theoretical model to the observed data, we subsequently add the unit-root world productivity shock and the stationary asymmetric (or foreign) technology shock to the business cycle component of foreign output in order to obtain the observed level of foreign GDP. This enables us to identify the stationary asymmetric technology shock, since the process for detrended foreign output is identified from the VAR and the process for the (unit root) world productivity is identified from this and the domestic quantities.

interest rate and inflation are weighted together across Sweden's 20 largest trading partners in 1991 using weights from the IMF.[7]

We include a large set of variables in the observed data vector, and match the following 15 variables: the GDP deflator, the real wage, consumption, investment, the real exchange rate, the short-run interest rate, hours worked, GDP, exports, imports, the consumer price index (CPI), the investment deflator, foreign output, foreign inflation and the foreign interest rate. As in Altig et al. (2003), the unit root technology shock induces a common stochastic trend in the real variables of the model. To make these variables stationary we use first differences and derive the state space representation for the following vector of observed variables

$$\tilde{Y}_t = \begin{array}{ccccccccc} [& \pi_t^d & \Delta\ln(W_t/P_t) & \Delta\ln C_t & \Delta\ln I_t & \hat{x}_t & R_t & \hat{H}_t & \Delta\ln Y_t ... \\ & \Delta\ln\tilde{X}_t & \Delta\ln\tilde{M}_t & \pi_t^{cpi} & \pi_t^{def,i} & \Delta\ln Y_t^* & \pi_t^* & R_t^* &]'. \end{array} \quad (19)$$

The growth rates are computed as quarter to quarter log-differences, while the inflation and interest rate series are measured as annualized quarterly rates. It should be noted that the stationary variables $\hat{x}_t$ and $\hat{H}_t$ are measured as deviations around the mean, i.e. $\hat{x}_t = (x_t - x)/x$ and $\hat{H}_t = (H_t - H)/H$, respectively. We choose to work with per capita hours worked, rather than total hours worked, because this is the object that appears in most general equilibrium business cycle models.[8]

In comparison with prior literature, such as for example Justiniano and Preston (2004) and Lubik and Schorfheide (2005), we have chosen to work with a large number of variables because we believe that it facilitate identification of the parameters and shocks we estimate. We estimate 13 structural shocks of which 8 follow AR(1) processes and 5 that are assumed to be identically independently distributed. In addition to these there are eight shocks provided by the exogenous (pre-estimated) fiscal and foreign VARs, whose parameters are kept fixed at their posterior mean estimates throughout the estimation of the DSGE model parameters. The shocks enter in such a way that there is no stochastic singularity in the likelihood function.[9] To compute the likelihood function, the reduced form solution of the model is transformed into a state-space representation mapping the unobserved state variables into the observed data. We apply the Kalman filter to calculate the likelihood function of the observed variables, where the period 1980Q1-1985Q4 is used to form a prior on the unobserved state variables in 1985Q4 and the period 1986Q1-2004Q4 for inference.

We choose to calibrate those parameters which we think are weakly identified by the variables that we include in the vector of observed data. These parameters are mostly related to the steady-state values of the observed variables (i.e., the great ratios: C/Y, I/Y and G/Y), see

[7] The shares of import and export to output are increasing from about 0.25 to 0.40 and from 0.21 to 0.50 respectively during the sample period. In the model, import and export are however assumed to grow at the same rate as output. Hence, we decided to remove the excess trend in import and export in the data, to make the export and import shares stationary. For all other variables we use the actual series (seasonally adjusted with the X12-method except the variables in the GDP identity which were seasonally adjusted by Statistics Sweden).

[8] We used working age population to generate hours per capita. See Christiano, Eichenbaum and Vigfusson (2003) for a discussion on using per capita adjusted hours versus non-adjusted hours.

[9] Even if there is no stochastic singularity in the model we include measurement errrors in the 12 domestic variables, since we know that the data series used are not perfectly measured and at best only approximations of the 'true' series. In particular it was hard to remove the seasonal variation in the series, and there are still spikes in for example the inflation series, perhaps due to changes in the collection of the data. The variance of the (uncorrelated) measurement errors is set to 0 for the foreign variables and the domestic interest rate, 0.1 percent for the real wage, consumption and output, and 0.2 percent for all the variables. This implies that the fundamental shocks explain about 90-95% of the variation in most of the variables. It should also be noted that the measurement errors are capturing some of the high frequency movements in the data and not business cycle fluctuations.

Table 1. An alternative approach could be to include these parameters in the estimation. However, such a strategy would require a different set of variables to ensure proper identification, and would yield similar results since these parameters would simply capture the sample mean of the great ratios.

The parameters we choose to estimate pertain mostly to the nominal and real frictions in the model as well as the exogenous shock processes. Table 2 shows the assumptions for the prior distribution of the estimated parameters. The location of the prior distribution of the 43 estimated parameters with no break in the monetary policy rule corresponds to a large extent to those in Adolfson et al. (2005a) on Euro area data, and are more thoroughly discussed in Adolfson et al. (2006).

The joint posterior distribution of the estimated parameters is obtained in two steps. First, the posterior mode and Hessian matrix evaluated at the mode is computed by standard numerical optimization routines. Second, the Hessian matrix is used in the Metropolis-Hastings algorithm to generate a sample from the posterior distribution (see Smets and Wouters (2003), and the references therein, for details). Table 2 reports the median estimates based on a sample of 500,000 post burn-in draws from the posterior distribution.

3. Maximum likelihood estimation on artificial data

In this section, we describe in detail how the parameter distributions have been generated from the artificial samples obtained out of the DSGE model. The following steps are conducted:

1. Solve the DSGE model using the calibrated parameters (see Table 1) and the posterior median of the estimated parameters (see Table 2).

2. Generate an artifical sample of length T by simulating the solved model $1000 + T$ periods initiated from the steady state. The first 1000 observations are discarded as burn-ins. The innovations in the shock series were drawn from the normal distribution, where we set the seed for each sample to $i = 1, ..., N$ where N is the number of artificial samples considered.[10]

3. The calibrated parameters in Table 1 and the size of the measurement errors are kept fixed at the 'true' values used to generate the artifical data. As a consequence, the ML estimation results will not reflect any uncertainty stemming from these parameters.

4. Given the artificial data (and the calibrated parameters), we estimate the parameters in Table 2 by maximizing the likelihood function using the same set of observable variables as on the actual data (see eq. 19). We use Chris Sims' optimizer CSMINWEL to perform the estimation.[11]

[10] An alternative to sampling from the normal distribution would be to bootstrap the innovations in the shock processes from the empirical distribution of the 2-sided estimates. But given that the purpose of the paper is to examine whether ML estimation can retrieve the true parameters used in the underlying data generating process, the bootstrapping approach is not appealing since the 2-sided estimates of the shock innovations are most likely heteroscedastic, autocorrelated and cross-correlated, which is at odds with the assumptions in the DSGE model.

[11] In the ML estimations, we impose the lower (b_l) and upper bounds (b_u) reported in the last two columns in Table 2. In cases where the solution algorithm fail to solve the model, the log-likelihood function is set to $-200,000$. We use the following smooth mapping function $p_{\text{mod}} = b_u - \frac{b_u - b_l}{1+e^{p_{\text{opt}}}}$ between the model parameters (p_{mod}) and the parameters that we optimize over (p_{opt}). Notice that p_{mod} converges to b_u when p_{opt} approaches ∞, and that p_{mod} converges to b_l when p_{opt} approaches ∞.

5. We store the resulting parameter estimates along with the likelihood information, inverse hessian, seed number used to generate the sample, and convergence diagnostics.

6. We repeat Step 1 to 6 a sufficiently large number of times to obtain a distribution that is stable. In practice it took between $1,000$ and $1,500$ estimations to obtain approximate convergence in mean and variance in the distribution for each estimated parameter.

One important difference w.r.t. how the model was estimated on actual data in the previous Section, is that we do not include measurement errors in the estimation. Also, we fix the parameters of the exogenous foreign and fiscal policy VARs at their true values throughout the analysis. The only reason for this is to simplify the interpretation of the results, and focus on the key model parameters in Table 2. But we have also conducted ML estimations when we add measurement errors to the artificial model data in line with how they were calibrated on Swedish data. In this case, we also used artifical data to estimate VAR(4) and VAR(2) models for the foreign and fiscal variables respectively (where the foreign output gap variable and government expenditure series are computed using the HP-filter) instead of treating the VARs for the foreign and fiscal variables to be known and kept fixed at their true values in each estimation. This alternative approach of incorporating measurement errors and estimated fiscal and foreign VARs did not change the bias and consistency properties of the ML estimation results reported in the figures and tables below, but it widened the dispersion in the parameter distributions somewhat.[12]

We consider two sample size values. As a benchmark, we set $T = 100$, which is equivalent to the size of our actual sample. In order to examine potential small sample problems and check consistency of the ML estimator we also generate distributions when we set $T = 400$. The results in the tables and figures below are based on the convergent estimations only, but we will provide the fraction of simulations that did not converge. This choice reflects our belief that the econometrician would not be satisfied with an estimation that led to a non-convergent estimation, and would redo the estimation by perturbing the starting values of the optimization until a satisfactory convergence was found. However, we instead decided to draw a new sample and continue. As will be clear later, however, this choice do not seem to be critical.

We initiate the estimations with the true parameter values, but we will also report results for alternative starting values. This will be done in two ways. First, we randomize the starting values by sampling parameters from the prior distribution. Second, we also conduct some estimations when we initialize the estimation with the prior mode values of the parameters instead of the posterior median.

We will also report results when the model is estimated on a subset of variables, namely the 7 “domestic” variables matched by Smets and Wouters (2003). That is, we drop all “open” economy variables such as for example the real exchange rate and exports and imports. Finally, to learn more about the curvature of the likelihood function, we compute a distribution of estimates based on only one artificial sample, using different starting values in the estimations by sampling these from the prior distribution. This exercise has two interesting aspects. First, in the best of worlds, one would hope that these estimations always converge to the same log-likelihood function value regardless of starting value. Second, even if the ML estimator does not converge to the same likelihood in all estimations when we sample intial starting values from the prior distributions, it should at least be the case that the estimations on a given sample all produce identical estimates at every time the estimation converges to roughly the same likelihood function. If the marginal distributions of the parameters have not collapsed although the ML

[12] These results are available in Appendix A.1. See in particular Figures A1.a-c.

estimations have returned to the same log likelihood, it is a strong sign that some parameters are weakly identified.

4. Monte-Carlo simulation results

In this section we provide the results of the Monte-Carlo simulations. We report statistics of the simulated distributions in Tables 3 and 4, and in Figures 1-6, we report kernel density estimates of the various parameter distributions. In order to impose that all kernel density estimates in the figures are within the plausible range for the parameters (e.g. between 0 and 1 for the Calvo parameters), the kernel density estimates are computed in the unbounded parameter space in which the optimizer actually works.[13] The kernel density estimates are then transformed to the bounded parameter space and depicted in the figures below.

4.1. Benchmark results: initializing estimations from the true parameters

In Table 3, we report the results when initializing the optimizations from the true parameter values. Results for two sample sizes are reported, $T = 100$ and $T = 400$. As can be seen from the table, almost every parameter's mean and median estimates equal the true parameter already for a sample size of $T = 100$. So the ML estimator appears to be an unbiased estimator for almost every parameter in the model. Two important exceptions are the coefficients in the policy rule, r_π and r_y, which both have mean estimates that are much higher than their true values. However, the medians for the two parameters are of the right magnitude, suggesting that the high mean values are driven by outliers in the distribution. Given the specification of the instrument rule, where ρ_R multiplies the coefficients in the policy rule (see eq. 15), it is perhaps not surprising that the distributions for these two parameters can be skewed to the left. In samples when ρ_R becomes high, r_π and r_y can easily end up at very high values. The fourth column of Table 3 shows the standard deviation of the simulated distributions, and not surprisingly the standard deviations are very high for these two parameters. The standard deviations are also relatively high for the investment adjustment cost parameter, $\tilde{S}''$, suggesting that also this parameter is sometimes driven to a very high number. In addition to the standard deviations of the resulting parameter distribution, Table 3 reports in the fifth column the median standard deviation of the estimates in each sample using the inverse Hessian matrix.[14] The median standard deviations for each of the ML estimates are generally somewhat smaller than the standard deviations in the parameter distributions, and in this sense they appear to underestimate the uncertainty about the ML parameters. However, the median standard deviations based on the inverse Hessian are at least accurate in the sense that they convey more uncertainty for those parameters where the standard deviations of the simulated parameter distributions are high. But for the parameters that are skewed, like $\tilde{S}''$ and r_π and r_y the median standard deviations clearly underestimate the true degree of uncertainty.

Turning to the results for $T = 400$, we see that the mean and median parameter estimates are getting more similar in general, and for $\tilde{S}''$ and r_π and r_y in particular. Both the mean and

[13] See the mapping function in Footnote 11.

[14] The inverse Hessian has full rank and is positive definite with the exception of a few simulations (22 out of the convergent 1474) in the benchmark estimations. When a number of variables are excluded in the information set that is used to estimate the model, the number of inverse Hessians that do not have full rank and are positive definite increases sharply. Notice that since the parameter optimizations are done in the transformed parameter space (see Footnote 11), the standard deviations are computed by assuming normality of the estimated parameters in each optimization and using the inverse Hessian and point estimates in the unbounded space to form a distribution in the bounded parameter space, for which the covariance matrix is computed.

median is now also very similar to the true parameter values. In addition, it is clear that the distributions start to collapse around the true values as the standard deviations of the marginal distributions have been reduced by at least a factor of 2, and in some cases even more. The median standard deviations of the estimates are also more accurate for this sample size, but there is still a tendency of underestimating the uncertainty in the parameter distribution.

In Figures 1a-1c, we complement the information in the table by plotting the kernel density estimates of the marginal parameter distrubutions. The figure confirms the picture in Table 3 and shows that the distributions for $\tilde{S}''$ and r_π and r_y are clearly skewed to the left. But the figure also makes it very clear that this set of data suffices for identification of the true parameters in the notion of Rothenberg (1971): as the sample size increases, the parameter distributions start to collapse around the true parameters. So conditional on this number of variables and the parameter we seek to estimate, the ML estimator appear to be consistent.[15]

The results sofar describe a somewhat different picture than the one by Canova and Sala (2006), who question the ability to achieve identification in DSGE models. However, although the marginal distributions are satisfactory from a frequentistic perspective, the arguments brought to the table by Canova and Sala (2006) are partly supported by computing pairwise correlations between parameters, and graphing the bivariate distributions. In Figure 2, we show all the pairwise parameter combinations with correlation above 0.5. In the graph, we also include the parameter correlation coefficient. The figure gives clear support for the idea that there is a large but not perfect degree of substitutability between some of the parameters in the model. Some parameter combinations imply a certain degree of partial identification. In particular, Figure 2 suggests that this problem pertains to three sets of parameters. First, we see that many of the parameters in the policy rule are highly correlated with each other. For example, there is a clear positive and non-linear relationship between ρ_R and $\{r_\pi, r_y\}$ and negative correlation between ρ_R and r_x, which is not surprising given that these coefficients enter multiplicative in the Taylor rule (15). The second set of parameters which exhibit a high degree of substitutability is some of the persistence and standard deviation parameters of the shock processes. This feature pertains to the unobserved AR(1) shock process for the unit root technology shock ($\mu_{z,t}$), the investment specific technology shock (Υ_t), the exchange risk-premium shock ($\tilde{\phi}_t$) and the labor supply shock (ζ_t^h). There is quite naturally a negative correlation between these parameters, suggesting that the ML estimator has difficulties in distinguishing whether the combination of either high persistence/low variance of the innovations or low persistence/high variance of the innovations is most plausible for these parameters. Again, this result is not surprising. The third set of parameters which exhibit a high degree of linear dependence is a set of parameters pertaining to the open economy aspects of the model. In particular, some of the markup parameters on imported consumption and investment goods, and the elasticity of substitution between domestically and imported investment goods are highly correlated. Especially the pairwise correlation between λ_{mi} and η_i is very high, suggesting that one of them could have been calibrated and not been included in the estimation.[16] Perhaps more surprisingly, there is a clear

[15] In addition to matching quantity variables in first differences, we have also studied the properties of the ML estimator when including the true co-integrating vectors in the set of observed quantity variables. These results are reported in Appendix A.2. The results show that there are only very small efficiency gains for ML estimation when matching the co-integrating vectors as opposed to matching the quantities and the real wage in first differences.

[16] However, below we will argue that this is not the case in a more global sense. The high degree of linear dependence between the markup and import/export elasticity parameter is only locally in the parameter space, e.g. for $\lambda_{m,i}$ between 1.1 and 1.2. They are well identified more globally (i.e. the data is really informative that $\lambda_{m,i}$ should be in the 1.1-1.2 range and not for example 1.6. **[Do pairwise contour plots to show this feature of the model (for a given sample?)**]

positive relationship between the standard deviation of the risk premium shock (σ_ϕ) and the endogenous risk premium coefficient ϕ_s. Finally, there are a number of parameters pertaining to exports that are highly correlated. This is not a surprising finding, however, because the only variable that is directed at pinning down the parameters pertaining to the export sector is the export quantity variable. Because of the local currency pricing assumption for the exporting firms, we have not been able to include an export price variable as observable in the estimation of the model. If this was possible, it is quite clear that the problems pertaining to the export parameters would be moderated.

4.2. Initiating estimations by sampling from the prior distributions

In the previous subsection, a key feature of the estimations was that they were initiated from the true parameter values. This is a clear advantage for the ML estimator in a large model, and perhaps an unrealistic assumption in practice. If we always start out with the true parameters - why bother doing estimations at all? In addition, if the multidimensional likelihood surface is characterized by many local maximas, there is the possibility that the favorable results in the previous subsection was driven by the very good guesses that intialized the estimations. In this subsection we therefore relax this assumption and instead initialize the optimizations by sampling from the prior distribution in Table 2 that were used to estimate the model on actual data. We construct a joint distribution of the parameters in the following way. First, we make $30,000$ draws from the prior distribution. Then we compute the 2.5 and 97.5th percentiles for each parameter in this distribution, and select all draws in the joint distribution that simultaneously are within the 2.5th and 97.5th percentiles. This procedure gives a distribution of starting values that in some cases differ substantially from the true parameter values because some of the priors in Table 2 are relatively uninformative (in particular the priors for the standard deviations of the shock processes). Therefore, we expect a larger fraction of non-convergent estimations in the ML estimations. This is also confirmed in the simulations, where only about 700 out of the $N = 1,500$ simulations converge. However, this number of convergent simulations is sufficient to obtain a fairly stable distribution (as argued below).

In Table 4, we report the mean, median and standard deviation of the distributions when starting out the optimizations from the prior distribution and when starting out from the true parameter values for the same set of convergent samples. So in both cases, only results for the same samples are reported in order to be able to make an accurate comparison. The results in the Table 4 can also be compared to the results in Table 3 for $T = 100$, which were based on nearly all the $1,500$ samples. From this comparison, it is clear that the distributions are roughly the same, so the results in Table 4 are not plauged by having to few convergent samples.

But an even more interesting comparison is to compare the distribution from "true initialization" against the distribution resulting from "prior initialization". By doing so, it is clear from Table 4 that there are very few differences between the two distributions, so the intial guess does not seem to be imperative for the performance of the ML estimator. The exceptions are not surprisingly the three parameters $\tilde{S}''$ and r_π and r_y which have higher and distorted means, but their median estimates are still accurate, which is a clear indication that their means are driven by outliers. So even if we start out from non-true starting values, the ML estimator is unbiased in most cases if we consider the median estimates. There is one problem with this interpretation though, namely that it might be the case that the only estimations that have converged are associated with starting values very close to the true parameters. And if this is true, the results in Table 4 cannot be taken as basis for declaring success.

For this reason, we compare the distribution resulting from "true initialization" (solid black) against the distribution resulting from "prior initialization" (dashed black) along with the actual starting value distribution (dotted line) in Figure 3. From the figure, it is clear that the prior distributions for the 700 convergent estimations we used are clearly off relative to the true parameter values in line with the priors used on actual data (see Table 2). So it is not the case that the ML estimator is able to find the optimum only for starting values sampled from the prior that are nearly identical to the true parameters. The optimizations can be initiated with parameters that are far off optimum and convergence can still be achieved. Another striking feature of Figures 3a-3b is that the distributions are very similar in most cases. Basically the 95-percent bands are identical, but the tails are fatter in the distribution resulting from the prior initialization, suggesting that the ML estimator sometimes sets off for implausible parameter regions. This affects the standard deviations of the distributions as can be verified in Table 4 (should consider including 2.5th and 97.5th percentiles in Table 4).

To sum up, the main conclusion is that the satisfactory performance of the ML estimator still holds, even if the econometrician does not have a perfect guess of the starting value of the parameters.

4.3. Sampling initial values from the prior distributions: One artifical data set

To complement the analysis above, and to get a deeper understanding of which parameters are associated with weak identification, we take a given dataset (i.e. the dataset that is generated when the seed is set to 1). For this dataset, we perform $1,500$ estimations where the starting values in the optimizations are sampled from the prior distribution as in the previous subsection. Out of the $1,500$ estimations, 638 converged according to the criterias of csminwel. Out of these 638 convergent optimizations, 545 optimizations converged to interior solutions (i.e. did not hit neither the upper nor the lower bounds) with plausible likelihood values. To study how informative the likelihood function is about the parameters, we use all simulations that have converged to the same likelihood out.of the 545 remaining samples In practice, we took the best likelihood function value and accepted samples with up to 0.1 units lower log likelihood (the maximum log likelihood equals -1440.425 for this particular sample). By this procedure, we obtain 489 samples that have converged to very similar log-likelihood values. Now the interesting issue is: does this imply that the parameter estimates have converged to the same values as well?

In Figures 4a-4c, we plot the resulting parameter estimates as histograms, along with the kernel density estimates of the prior distributions that were used as starting values in the optimizations. As can be seen from the graphs, it is clear that some of the parameters are characterized by weak identification problems, in the sense that quite some variation in the parameter results in little variation in the log-likelihood function. Perhaps surprisingly, one of the most problematic parameters are ξ_w, the degree of nominal wage stickiness. But also other other parameters like r_π, r_y, the inflation target shock ($\sigma_{\bar{\pi}}$) and the persistense coefficient for the consumption preference shock (ρ_{ξ^c}) vary substantially. The results in Figures 4a-4c therefore complement the information contained in Figures 1a-1c and Figure 2, but also give a somewhat different perspective on identification. For instance, according to Figure 2, one would be tempted to draw the conclusion that $\lambda_{m,i}$ and $\lambda_{m,c}$ are not well identified. But according to Figure 4a, they are separately very well identified to a specific neighboorhood. But in this neighboorhood, they are very highly correlated and very weakly identified.

Our finding that a key parameter like ξ_w is weakly identified raises the issue of what feature in the DSGE model that leads to this finding.[17] The parameterization of the data generating

[17] From Appendix A.2, it is clear that imposing the true co-integrating vector for the real wage in the estimations, i.e. matching $\ln(W_t/P_t) - \ln Y_t$ instead of $\Delta \ln(W_t/P_t)$ does not mitigate the problem with weak identification

process (i.e. the median estimates in Table 2) is characterized by a high degree of price stickiness and a high variance of the labor supply shocks. Our conjecture at this point, is that this drives the weak identification results for ξ_w, but this needs to be confirmed in simulations where we lower the variance of the labor supply shocks and the degree of nominal price stickiness in the model. [**Remains to be done. Do simulations where we shrink the sticky price parameters and the variance of the labor supply shocks.**]

4.4. Estimation on a subset of observable variables

In both subsections above, we used all the 15 variables in eq. (19) as observables when taking the model to the data. To understand how the performance of the ML estimator depends on the choice of observed variables, we now assume now that, for some reason, the econometrican only includes 7 variables when estimating the model. More specifically, we assume that the the following subset of variables in (19) is used:

$$\tilde{Y}_t^{subset} = \left[\begin{array}{ccccccc} \pi_t^d & \Delta \ln(W_t/P_t) & \Delta \ln C_t & \Delta \ln I_t & R_t & \hat{H}_t & \Delta \ln Y_t \end{array} \right]' . \tag{20}$$

The variables in (20) are the "closed economy" variables used by Smets and Wouters (2003), but for the sake of the argument we assume that the econometrician still tries to estimate all 43 parameters in Table 2.

For the sample size $T = 100$, we plot the resulting distributions (dashed line in Figures 5a-5c) based on (20) along with the distribution that is obtained when all 15 variables are used as observables (i.e. the benchmark results for $T = 100$ reported in Table 3 and Figures 1a-1c). In both cases, we initialize the estimations with the true starting values.

As can be seen from Figures 5a-5c, restricting the set of observable variables in (20), leads to substantially more dispersion in the parameter distributions. In particular, this is the case for the parameters related to the open economy aspects of the model. For instance, the uncertainty about $\xi_{m,c}$, $\xi_{m,i}$ and ξ_x as measured by the standard deviation in the simulated distributions are now substantially higher. It is also the case that the number of convergent estimations fall from $1,452$ to $1,147$, and around 160 times, the inverse Hessian fails the rank test in Matlab, suggesting that the DSGE model estimated on the variables in (20) only is close to rank deficient in many cases and therefore is on the borderline of being identified in the Rothenberg (1971) sense (i.e. suffer from a very stong degree of weak identification).

This exercise demonstrates that the econometrician needs to be very careful when selecting the number of variables in estimating the model. If classical estimation techniques are applied, it is imperative to think hard about the structure of the model and which variables that needs to be included in order to ensure identification of a given set of parameters in small samples.[18]

5. Concluding remarks

In this paper we have analyzed the properties of maximum likelihood in a state-of-the-art open economy new Keynesian DSGE model. Our analysis suggests that our open economy DSGE

for ξ_w.

[18] However, in Appendix A.3, we show results for $T = 1600$ and $T = 6400$ observations in each sample, for which the estimations are initiated from the prior mode in Table 2 instead. By doing so, we can learn if there is some information asymptotically to identify the parameter, and to what extent the results in Figures 5a-5c are driven by the fact that we start out from the true values in the optimizations. Perhaps surprising, the results show that the ML estimator is consistent even if only the variables in equation (20) are used, and that the optimizations converge to the true parameter values even in this case when the sample size is increased. So the results in Figures 5a-c are not driven by the fact that we start out from the true parameters in the optimizations.

model is identifiable in the notion of Rothenberg (1971): if an appropriate set of variables are used to estimate the DSGE model, the ML distributions collapse at the true parameters as the sample size is increased. In this sense, our results based on full information methods go against the limited information results in Canova and Sala (2006), who questions identification in the the new generation of DSGE models. However, the results in this paper also lends some support to the arguments in Canova and Sala (2006) regarding weak identification of some parameters.With weak identification we mean that quite some variation in certain of the parameters were consistent with only marginal changes in the likelihood. In the parameterization of the model we considered, one such parameter is the degree of nominal wage stickiness. As this is a key parameter in the new generation of DSGE models, we need to explore the reason for the weak idenfication pertaining to this parameter. Another feature of estimated DSGE models emphazied by Canova and Sala (2006) is that there appears to be a high but not perfect degree of substitutability between some of the parameters in the model.[**Write more about our findings in this regard.**]

We think our results warrants the use of more pre-checking and carefulness about which variables that needs to be included when bringing a model to the data. Basically, the econometrician that consider applying ML estimation needs to perform the sort of tests conducted in this paper to learn about the properties of the model before taking the model to actual data. In the case the econometrican considers a Bayesian approach, the issues discussed in this paper might not appear to be problematic, because the posterior for the parameters where the observed variables are not informative will equal the prior and thus inform the econometrician that some of the parameter are very weakly identified. However, we still think there are at least two distinct reasons why a Bayesian econometrician should pay attention to the issues analyzed in this paper. First, a Bayesian investigator would certainly be interested to know what set of variables he needs to include in the set of variables used in estimation to be able to update his priors from the observed data sample in an efficient way. Second, the Bayesian econometrician needs to consider that the estimation results can be driven by the fact that (some of) his priors are more informative than others. For instance, the estimation results for an exogenous shock process can be (at least in a local neighbourhood) driven by the fact that the prior for the persistence coefficient of a shock (typically a rather informative beta prior) is more informative relative to the standard deviation of the innovation for the shock process (typically a rather uninformative inverse gamma prior).

Finally, there are at least three interesting aspects that certainly deserves future attention. First, and perhaps most importantly, Rubio and Villaverde (2005) compares maximum likelihood estimation of a real business cycle model and argues that estimations based on a non-linear (i.e. second-order) approximation are much more informative about the underlying parameters as opposed to estimations when the underlying DSGE model is log-linearized. Therefore, an interesting extension of the work here would be to examine to what extent the properties of maximum likelihood would be enhanced by working with the second-order approximations instead of a log-linearized represention of the model.

Second, there is also more work to be done in understanding the role of various real and nominal frictions for acheiving identification, and why a key parameter like the degree of nominal wage stickiness suffer from weak identification. One potential explanation for the latter issue can be the relatively high variance of the wage markup/labor supply shock and the high degree of sticky prices in the datagenerating process.[**Write about our findings.**]

Third, the issue about smaller models vs. larger models and larger sets of observed variables is an interesting extension as well. For instance, one could estimate the closed economy version of the model above and see if we learn more efficiently about the closed economy parameters in a closed economy setting compared to the open economy setting.

[**To be continued.**]

References

Adolfson, Malin, Stefan Laséen, Jesper Lindé and Mattias Villani (2006), "Evaluating An Estimated New Keynesian Small Open Economy Model", *Journal of Economic Dynamics and Control*, forthcoming.

Adolfson, Malin, Stefan Laséen, Jesper Lindé and Mattias Villani (2007), "Bayesian Estimation of an Open Economy DSGE Model with Incomplete Pass-Through", *Journal of International Economics*, 72(2), pp. 481-511.

Altig, David, Lawrence Christiano, Martin Eichenbaum and Jesper Lindé (2003), "The Role of Monetary Policy in the Propagation of Technology Shocks", manuscript, Northwestern University.

Anderson, Gary and George Moore (1985), "A Linear Algebraic Procedure for Solving Linear Perfect Foresight Models", *Economics Letters* 17(3), 247-252.

Canova, Fabio and Luca Sala (2006), "Back to Square One: Identification Issues in DSGE Models", Working Paper No. 303, IGIER Università Bocconi.

Cristadoro R., A. Gerali, S. Neri and P. Pisani, (2007), "Exchange rate volatility and disconnect: An empirical investigation", working paper, Banca d'Italia

Christiano, Lawrence, Martin Eichenbaum and Charles Evans (2005), "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy", *Journal of Political Economy* 113(1), 1-45.

Christiano, Lawrence, Martin Eichenbaum and Robert Vigfusson (2003), "What Happens After a Technology Shock?", NBER Working Paper, No. 9819.

Christoffel, K., G. Coenen and A. Warne (2007), "The New Area-Wide Model (NAWM) for the Euro Area: Specification, Estimation Results and Properties", June draft, European Central Bank.

Del Negro, Marco, Frank Schorfheide, Franks Smets and Rafael Wouters (2007), "On the Fit and Forecasting Performance of New Keynesian Models", *Journal of Business and Economic Statistics*, 25(2), pp. 123-162.

Duarte, Margarida and Alan Stockman (2005), "Rational Speculation and Exchange Rates", *Journal of Monetary Economics*, 52, 3-29.

Erceg, Christopher, Luca Guerrieri, and Christopher Gust (2006), "SIGMA: A New Open Economy Model for Policy Analysis", *Journal of International Central Banking* 2 (1), 1-50.

Erceg, Christopher, Dale Henderson and Andrew Levin (2000), "Optimal Monetary Policy with Staggered Wage and Price Contracts", *Journal of Monetary Economics* 46(2), 281-313.

Fama, Eugene, (1984), "Forward and Spot Exchange Rates", *Journal of Monetary Economics*, 14, 319–338.

Justiniano, Alejandro and Bruce Preston (2004), "Small Open Economy DSGE Models: Specification, Estimation and Model Fit", Manuscript, Columbia University.

Lubik, Thomas and Frank Schorfheide (2005), "A Bayesian Look at New Open Economy Macroeconomics", in eds. Gertler, Mark. and Kenneth Rogoff, *NBER Macroeconomics Annual.*

Lundvik, Petter (1992), "Foreign Demand and Domestic Business Cycles: Sweden 1891-1987", Chapter 3 in *Business Cycles and Growth*, Monograph Series No. 22, Institute for International Economic Studies, Stockholm University.

Pesenti, P., (2003) "The Global Economy Model (GEM): Theoretical framework", International Monetary Fund Working Paper.

Rabanal, Pau and Vincente Tuesta (2005), "Euro-Dollar Real Exchange Rate Dynamics in an Estimated Two-Country Model:

What is Important and What is Not", IMF Working Paper No. .

Rothenberg, T., (1971), "Identification in Parametric Models", *Econometrica*, 39, pp. 577-591.

Rubio-Ramirez, Juan F. and Jesús Fernández Villaverde (2005), "Estimating Dynamic Equilibrium Economies: Linear Versus Nonlinear Likelihood", *Journal of Applied Econometrics*, 20, pp. 891–910.

Schmitt-Grohé Stephanie and Martín Uribe (2001), "Stabilization Policy and the Costs of Dollarization", *Journal of Money, Credit, and Banking* 33(2), 482-509.

Smets, Frank and Raf Wouters (2002), "Openness, Imperfect Exchange Rate Pass-Through and Monetary Policy", *Journal of Monetary Economics* 49(5), 913-940.

Smets, Frank and Raf Wouters (2003), "An Estimated Stochastic Dynamic General Equilibrium Model of the Euro Area", *Journal of the European Economic Association*, 1(5), 1123-1175.

Table 1: Calibrated parameters

Parameter	Description	Calibrated value
β	Households' discount factor	0.999
α	Capital share in production	0.25
η_c	Substitution elasticity between C_t^d and C_t^m	5
σ_a	Capital utilization cost parameter	$1,000,000$
μ	Money growth rate (quarterly rate)	1.010445
σ_L	Labor supply elasticity	1
δ	Depreciation rate	0.01
λ_w	Wage markup	1.05
ω_i	Share of imported investment goods	0.70
ω_c	Share of imported consumption goods	0.40
ν	Share of wage bill financed by loans	1
τ^y	Labor income tax rate	0.30
τ^c	Consumption tax rate	0.24
$\rho_{\bar{\pi}}$	Inflation target persistence	0.975
g_r	Government expenditures-output ratio	0.30

Table 2: Prior and posterior distributions

		Prior distribution			Posterior distribution	Bounds	
Parameter		type	mean	std. dev. / df	median	lower	upper
Calvo wages	ξ_w	beta	0.750	0.050	0.765	0.01	0.99
Calvo domestic prices	ξ_d	beta	0.750	0.050	0.825	0.01	0.99
Calvo import cons. prices	$\xi_{m,c}$	beta	0.750	0.050	0.900	0.01	0.99
Calvo import inv. prices	$\xi_{m,i}$	beta	0.750	0.050	0.939	0.01	0.99
Calvo export prices	ξ_x	beta	0.750	0.050	0.874	0.01	0.99
Indexation prices	κ_p	beta	0.500	0.150	0.227	0.01	0.99
Indexation wages	κ_w	beta	0.500	0.150	0.323	0.01	0.99
Investment adj. cost	$\tilde{S}''$	normal	7.694	1.500	8.584	0.1	100
Habit formation	b	beta	0.650	0.100	0.679	0.01	0.99
Markup domestic	λ_d	truncnormal	1.200	0.050	1.195	1.01	10
Subst. elasticity invest.	η_i	invgamma	1.500	4	2.715	0.01	20
Subst. elasticity foreign	η_f	invgamma	1.500	4	1.531	0.01	20
Markup imported cons.	$\lambda_{m,c}$	truncnormal	1.200	0.050	1.584	1.01	10
Markup.imported invest.	$\lambda_{m,i}$	truncnormal	1.200	0.050	1.134	1.01	10
Technology growth	μ_z	truncnormal	1.006	0.0005	1.005	1.0001	1.01
Risk premium	$\tilde{\phi}$	invgamma	0.010	2	0.050	0.0001	10
UIP modification	$\tilde{\phi}_s$	beta	0.500	0.15	0.606	0.0001	1
Unit root tech. shock persistence	ρ_{μ_z}	beta	0.850	0.100	0.845	0.01	0.9999
Stationary tech. shock persistence	ρ_ε	beta	0.850	0.100	0.925	0.01	0.9999
Invest. spec. tech shock persistence	ρ_Υ	beta	0.850	0.100	0.694	0.01	0.9999
Risk premium shock persistence	$\rho_{\tilde{\phi}}$	beta	0.850	0.100	0.684	0.01	0.9999
Consumption pref. shock persistence	ρ_{ζ_c}	beta	0.850	0.100	0.657	0.01	0.9999
Labour supply shock persistence	ρ_{ζ_h}	beta	0.850	0.100	0.270	0.01	0.9999
Asymmetric tech. shock persistence	$\rho_{\tilde{z}*}$	beta	0.850	0.100	0.964	0.01	0.9999
Unit root tech. shock std. dev.	σ_z	invgamma	0.200	2	0.133	0.01	10
Stationary tech. shock std. dev.	σ_ε	invgamma	0.700	2	0.668	0.01	10
Imp. cons. markup shock std. dev.	$\sigma_{\lambda_{m,c}}$	invgamma	1.000	2	1.126	0.01	400
Imp. invest. markup shock std. dev.	$\sigma_{\lambda_{m,i}}$	invgamma	1.000	2	1.134	0.01	400
Domestic markup shock std. dev.	σ_λ	invgamma	1.000	2	0.807	0.01	100
Invest. spec. tech. shock std. dev.	σ_Υ	invgamma	0.200	2	0.396	0.01	100
Risk premium shock std. dev.	$\sigma_{\tilde{\phi}}$	invgamma	0.050	2	0.793	0.01	10
Consumption pref. shock std. dev.	σ_{ζ_c}	invgamma	0.200	2	0.263	0.01	5
Labour supply shock std. dev.	σ_{ζ_h}	invgamma	1.000	2	0.386	0.01	15
Asymmetric tech. shock std. dev.	$\sigma_{\tilde{z}*}$	invgamma	0.400	2	0.188	0.01	2
Export markup shock std. dev.	σ_{λ_x}	invgamma	1.000	2	1.033	0.01	20
Monetary policy shock	$\sigma_{R,}$	invgamma	0.150	2	0.239	0.01	2
Inflation target shock	$\sigma_{\bar{\pi}^c}$	invgamma	0.050	2	0.157	0.01	1.5
Interest rate smoothing	ρ_R	beta	0.800	0.050	0.913	0.01	0.99
Inflation response	r_π	truncnormal	1.700	0.100	1.674	1.01	1000
Diff. infl response	$r_{\Delta\pi}$	normal	0.300	0.050	0.098	-0.5	5
Real exch. rate response	r_x	normal	0.000	0.050	-0.016	-5	5
Output response	r_y	normal	0.125	0.050	0.125	-0.5	5
Diff. output response	$r_{\Delta y}$	normal	0.063	0.050	0.178	-0.5	5

*Note: For the inverse gamma distribution the mode and the degrees of freedom are reported. Also, for the parameters $\lambda_d, \eta_i, \eta_f, \lambda_{m,c}, \lambda_{m,i}$ and μ_z the prior distributions are truncated at 1.

Table 3: Distribution results from different sample sizes using true starting values

Parameter		True estimates	100 observations: Mean of distribution	100 observations: Median of distribution	100 observations: Std. of distribution	100 observations: Median std. of inverse Hessians	400 observations: Mean of distribution	400 observations: Median of distribution	400 observations: Std. of distribution	400 observations: Median std. of inverse Hessians
Calvo wages	ξ_w	0.77	0.74	0.75	0.13	0.07	0.76	0.76	0.05	0.03
Calvo domestic prices	ξ_d	0.83	0.81	0.82	0.04	0.03	0.82	0.82	0.02	0.01
Calvo import cons. prices	$\xi_{m,c}$	0.90	0.90	0.90	0.02	0.01	0.90	0.90	0.01	0.01
Calvo import inv. prices	$\xi_{m,i}$	0.94	0.94	0.94	0.02	0.01	0.94	0.94	0.01	0.01
Calvo export prices	ξ_x	0.87	0.86	0.86	0.04	0.02	0.87	0.87	0.01	0.01
Indexation prices	κ	0.23	0.22	0.22	0.06	0.05	0.22	0.22	0.03	0.02
Indexation wages	κ_w	0.32	0.32	0.32	0.15	0.07	0.32	0.32	0.07	0.04
Investment adj. cost	$\tilde{S}''$	8.58	8.98	8.08	4.08	2.02	8.64	8.54	1.36	0.99
Habit formation	b	0.68	0.67	0.67	0.07	0.05	0.68	0.68	0.03	0.02
Markup domestic	λ_d	1.20	1.21	1.20	0.14	0.09	1.20	1.20	0.06	0.04
Subst. elasticity invest.	η_i	2.72	2.72	2.71	0.13	0.11	2.71	2.71	0.06	0.05
Subst. elasticity foreign	η_f	1.53	1.59	1.45	0.59	0.23	1.54	1.53	0.13	0.09
Markup imported cons.	$\lambda_{m,c}$	1.58	1.58	1.58	0.01	0.01	1.58	1.58	0.00	0.00
Markup.imported invest.	$\lambda_{m,i}$	1.13	1.14	1.13	0.02	0.02	1.13	1.13	0.01	0.01
Technology growth	μ_z	1.01	1.01	1.01	0.00	0.00	1.01	1.01	0.00	0.00
Risk premium	$\tilde{\phi}$	0.05	0.06	0.05	0.02	0.01	0.05	0.05	0.01	0.00
UIP modification	$\tilde{\phi}_s$	0.61	0.61	0.60	0.05	0.03	0.61	0.61	0.02	0.01
Unit root tech. persistance	ρ_{μ_z}	0.85	0.80	0.83	0.14	0.06	0.84	0.85	0.05	0.03
Stationary tech. persistance	ρ_{ε}	0.93	0.89	0.90	0.08	0.03	0.92	0.92	0.02	0.01
Invest. spec. tech. persist.	ρ_{Υ}	0.69	0.65	0.67	0.13	0.06	0.69	0.69	0.04	0.03
Risk premium persistence	$\rho_{\tilde{\phi}}$	0.68	0.65	0.65	0.11	0.06	0.68	0.68	0.04	0.03
Consumption pref. persist.	ρ_{ζ_c}	0.66	0.59	0.61	0.18	0.08	0.64	0.65	0.07	0.04
Labour supply persistance	ρ_{ζ_h}	0.27	0.26	0.26	0.13	0.07	0.27	0.27	0.06	0.04
Asymmetric tech. persist.	$\rho_{\tilde{z}^*}$	0.96	0.73	0.84	0.28	0.09	0.93	0.95	0.11	0.02
Unit root tech. shock	σ_{μ_z}	0.13	0.14	0.14	0.05	0.03	0.13	0.13	0.02	0.01
Stationary tech. shock	σ_{ε}	0.67	0.66	0.65	0.06	0.05	0.67	0.67	0.03	0.03
Imp. cons. markup shock	$\sigma_{\lambda_{m,c}}$	1.13	1.13	1.12	0.11	0.10	1.13	1.13	0.05	0.05
Imp. invest. markup shock	$\sigma_{\lambda_{m,i}}$	1.13	1.14	1.13	0.11	0.10	1.14	1.13	0.05	0.05
Domestic markup shock	σ_{λ_d}	0.81	0.82	0.82	0.08	0.08	0.81	0.81	0.04	0.04
Invest. spec. tech. shock	σ_{Υ}	0.40	0.42	0.41	0.09	0.06	0.40	0.40	0.03	0.03
Risk premium shock	$\sigma_{\tilde{\phi}}$	0.79	0.82	0.80	0.21	0.12	0.80	0.80	0.08	0.06
Consumption pref. shock	σ_{ζ_c}	0.26	0.27	0.27	0.05	0.04	0.27	0.26	0.02	0.02
Labour supply shock	σ_{ζ_h}	0.39	0.39	0.39	0.06	0.04	0.38	0.38	0.03	0.02
Asymmetric tech. shock	$\sigma_{\tilde{z}^*}$	0.19	1.13	1.09	0.41	0.21	1.04	1.03	0.11	0.08
Export markup shock	σ_{λ_x}	1.03	0.15	0.16	0.06	0.04	0.18	0.18	0.02	0.02
Monetary policy shock	σ_R	0.24	0.24	0.23	0.02	0.02	0.24	0.24	0.01	0.01
Inflation target shock	$\sigma_{\bar{\pi}^c}$	0.16	0.14	0.14	0.10	0.04	0.16	0.16	0.03	0.02
Interest rate smoothing	ρ_R	0.91	0.91	0.91	0.05	0.03	0.91	0.91	0.02	0.02
Inflation response	r_{π}	1.67	3.80	1.59	5.08	2.70	2.06	1.65	1.59	0.61
Diff. infl response	$r_{\Delta\pi}$	0.10	0.11	0.10	0.04	0.03	0.10	0.10	0.02	0.01
Real exch. rate response	r_x	-0.02	-0.07	-0.02	0.15	0.02	-0.03	-0.02	0.04	0.01
Output response	r_y	0.13	0.35	0.13	0.63	0.07	0.17	0.12	0.18	0.04
Diff. output response	$r_{\Delta y}$	0.18	0.19	0.18	0.05	0.03	0.18	0.18	0.02	0.02

Note: Out of the 1500 estimations for the small sample (100 obs.), the results above is based on 1452 convergent estimations. Out of the 1000 estimations for the large sample (400 obs.), the results above is based on 999 convergent estimations.

Table 4: Distribution results from different starting values (no. of observations = 100)

Parameter		True estimates	Starting from true values			Starting from a distribution of prior values		
			Mean of distribution	Median of distribution	Std. of distribution	Mean of distribution	Median of distribution	Std. of distribution
Calvo wages	ξ_w	0.77	0.74	0.75	0.13	0.73	0.74	0.16
Calvo domestic prices	ξ_d	0.83	0.81	0.82	0.04	0.81	0.82	0.06
Calvo import cons. prices	$\xi_{m,c}$	0.90	0.90	0.90	0.02	0.90	0.90	0.03
Calvo import inv. prices	$\xi_{m,i}$	0.94	0.94	0.94	0.02	0.93	0.94	0.04
Calvo export prices	ξ_x	0.87	0.86	0.86	0.04	0.86	0.86	0.05
Indexation prices	κ	0.23	0.22	0.22	0.06	0.23	0.22	0.11
Indexation wages	κ_w	0.32	0.32	0.32	0.15	0.33	0.32	0.18
Investment adj. cost	$\tilde{S}''$	8.58	8.98	8.08	4.08	10.04	8.54	7.85
Habit formation	b	0.68	0.67	0.67	0.07	0.68	0.68	0.08
Markup domestic	λ_d	1.20	1.21	1.20	0.14	1.35	1.20	1.10
Subst. elasticity invest.	η_i	2.72	2.72	2.71	0.13	2.72	2.71	0.18
Subst. elasticity foreign	η_f	1.53	1.59	1.45	0.59	1.60	1.43	0.94
Markup imported cons.	$\lambda_{m,c}$	1.58	1.58	1.58	0.01	1.58	1.58	0.01
Markup.imported invest.	$\lambda_{m,i}$	1.13	1.14	1.13	0.02	1.14	1.14	0.03
Technology growth	μ_z	1.01	1.01	1.01	0.00	1.01	1.01	0.00
Risk premium	$\tilde{\phi}$	0.05	0.06	0.05	0.02	0.07	0.05	0.11
UIP modification	$\tilde{\phi}_s$	0.61	0.61	0.60	0.05	0.60	0.60	0.09
Unit root tech. persistance	ρ_{μ_z}	0.85	0.80	0.83	0.14	0.81	0.85	0.14
Stationary tech. persistance	ρ_{ε}	0.93	0.89	0.90	0.08	0.88	0.90	0.09
Invest. spec. tech. persist.	ρ_{Υ}	0.69	0.65	0.67	0.13	0.66	0.67	0.14
Risk premium persistence	$\rho_{\tilde{\phi}}$	0.68	0.65	0.65	0.11	0.66	0.66	0.12
Consumption pref. persist.	ρ_{ζ_c}	0.66	0.59	0.61	0.18	0.60	0.62	0.20
Labour supply persistance	ρ_{ζ_h}	0.27	0.26	0.26	0.13	0.27	0.26	0.16
Asymmetric tech. persist.	$\rho_{\tilde{z}*}$	0.96	0.73	0.84	0.28	0.73	0.83	0.28
Unit root tech. shock	σ_{μ_z}	0.13	0.14	0.14	0.05	0.14	0.13	0.08
Stationary tech. shock	σ_{ε}	0.67	0.66	0.65	0.06	0.66	0.65	0.13
Imp. cons. markup shock	$\sigma_{\lambda_{m,c}}$	1.13	1.13	1.12	0.11	1.24	1.12	2.02
Imp. invest. markup shock	$\sigma_{\lambda_{m,i}}$	1.13	1.14	1.13	0.11	1.39	1.13	6.64
Domestic markup shock	σ_{λ_d}	0.81	0.82	0.82	0.08	0.87	0.82	1.14
Invest. spec. tech. shock	σ_{Υ}	0.40	0.42	0.41	0.09	0.42	0.41	0.09
Risk premium shock	$\sigma_{\tilde{\phi}}$	0.79	0.82	0.80	0.21	0.83	0.80	0.46
Consumption pref. shock	σ_{ζ_c}	0.26	0.27	0.27	0.05	0.28	0.27	0.19
Labour supply shock	σ_{ζ_h}	0.39	0.39	0.39	0.06	0.42	0.39	0.58
Asymmetric tech. shock	$\sigma_{\tilde{z}*}$	0.19	1.13	1.09	0.41	0.16	0.16	0.12
Export markup shock	σ_{λ_x}	1.03	0.15	0.16	0.06	1.16	1.10	0.45
Monetary policy shock	σ_R	0.24	0.24	0.23	0.02	0.25	0.24	0.11
Inflation target shock	$\sigma_{\bar{\pi}^c}$	0.16	0.14	0.14	0.10	0.16	0.14	0.14
Interest rate smoothing	ρ_R	0.91	0.91	0.91	0.05	0.91	0.91	0.05
Inflation response	r_{π}	1.67	3.80	1.59	5.08	4.80	1.50	13.47
Diff. infl response	$r_{\Delta\pi}$	0.10	0.11	0.10	0.04	0.12	0.10	0.24
Real exch. rate response	r_x	-0.02	-0.07	-0.02	0.15	-0.05	-0.01	0.56
Output response	r_y	0.13	0.35	0.13	0.63	0.41	0.11	0.83
Diff. output response	$r_{\Delta y}$	0.18	0.19	0.18	0.05	0.20	0.18	0.32

Note: Out of the 1500 estimations, the results above are based on 688 commonly convergent estimations.

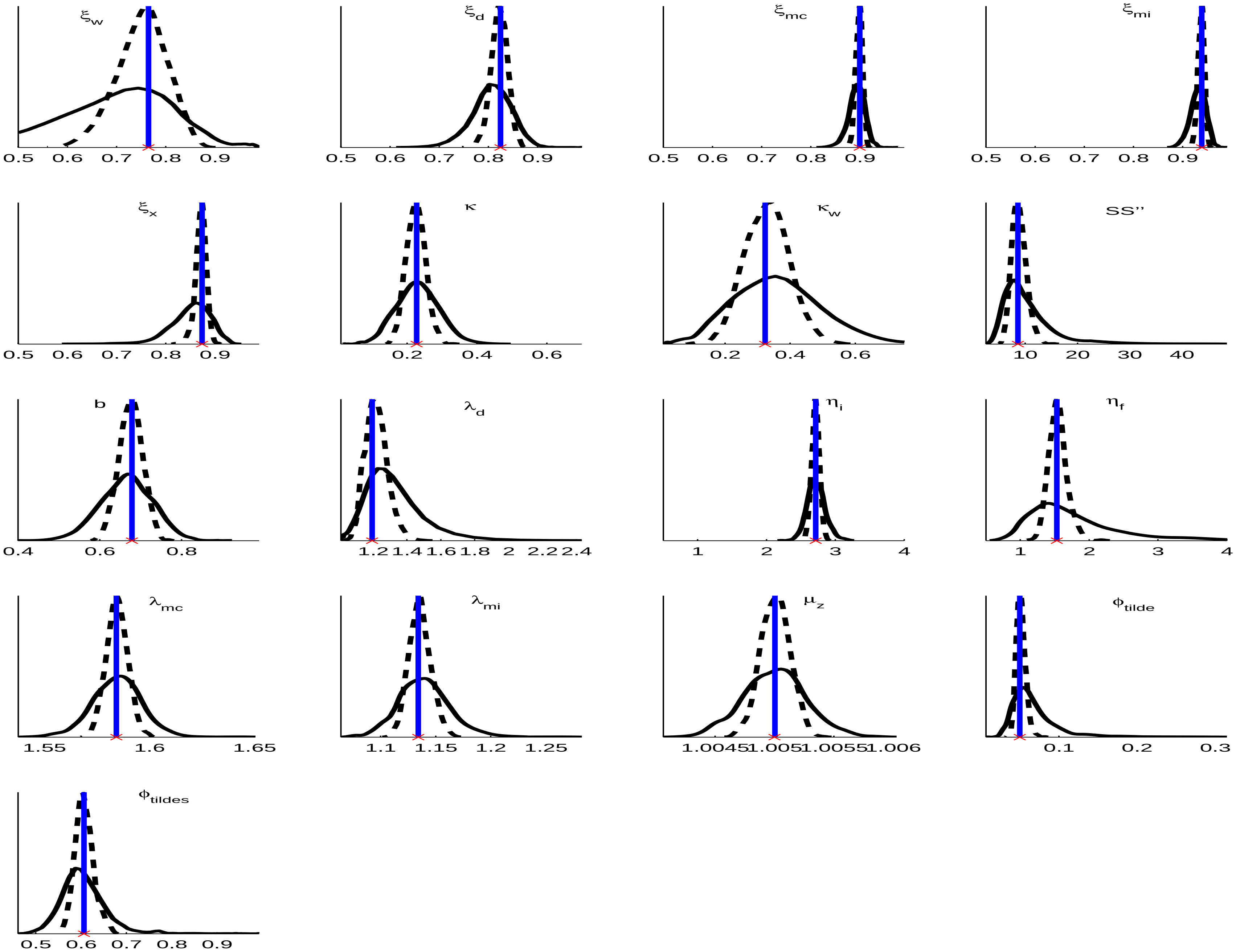

Figure 1a: Kernel density estimates of the small sample distribution for the estimates of the deep model parameters. The solid line show the parameter distribution for $N = 100$, and the dashed line shows the distribution for $N = 400$ observations. The vertical bar shows true parameter value, and the cross on the x-axis denotes the starting value in the optimizations.

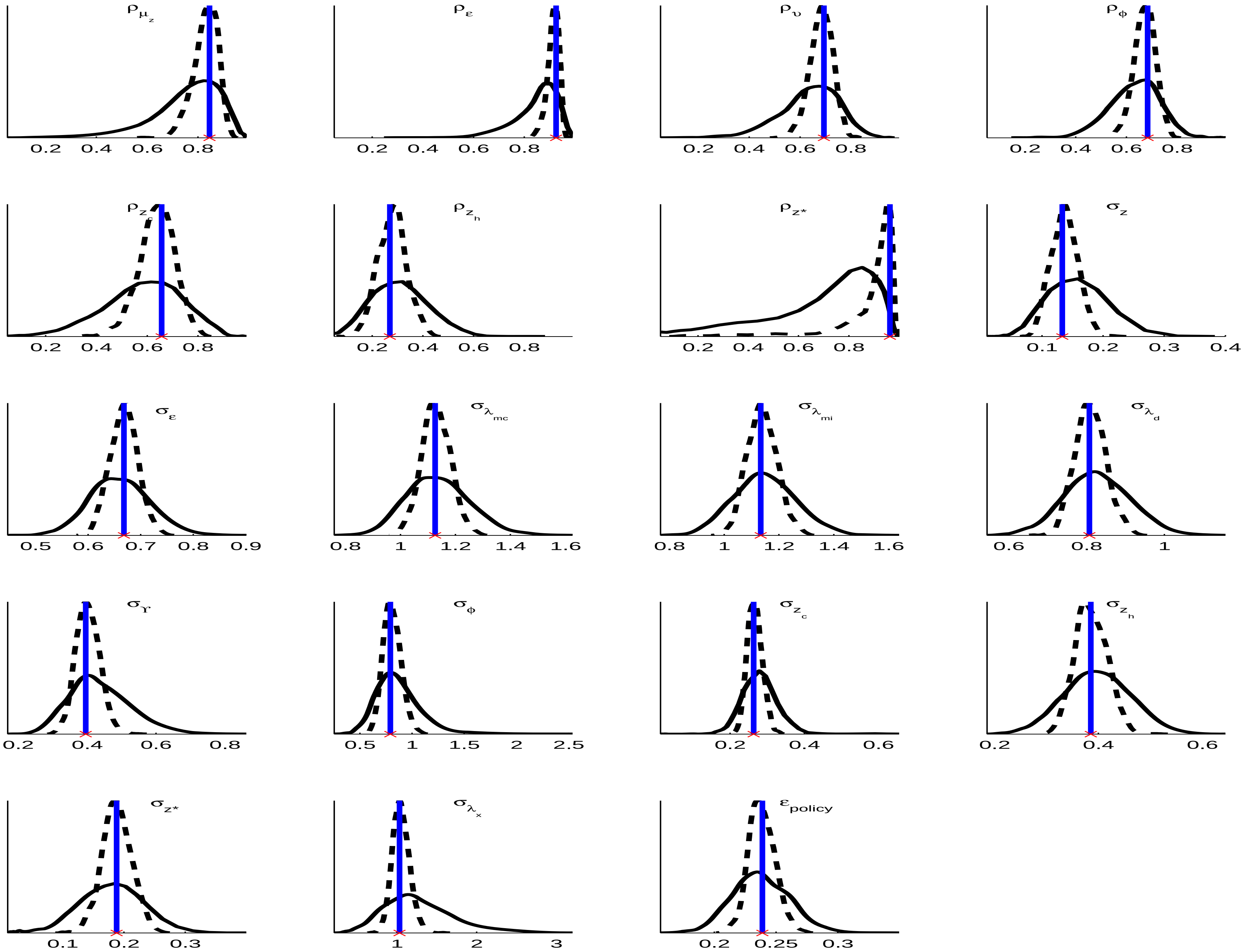

Figure 1b: Kernel density estimates of the small sample distribution for the estimates of the shock parameters. The solid line shows the parameter distribution for $N = 100$, and the dashed line shows the distribution for $N = 400$ observations. The vertical bar shows true parameter value, and the cross on the x-axis denotes the starting value in the optimizations.

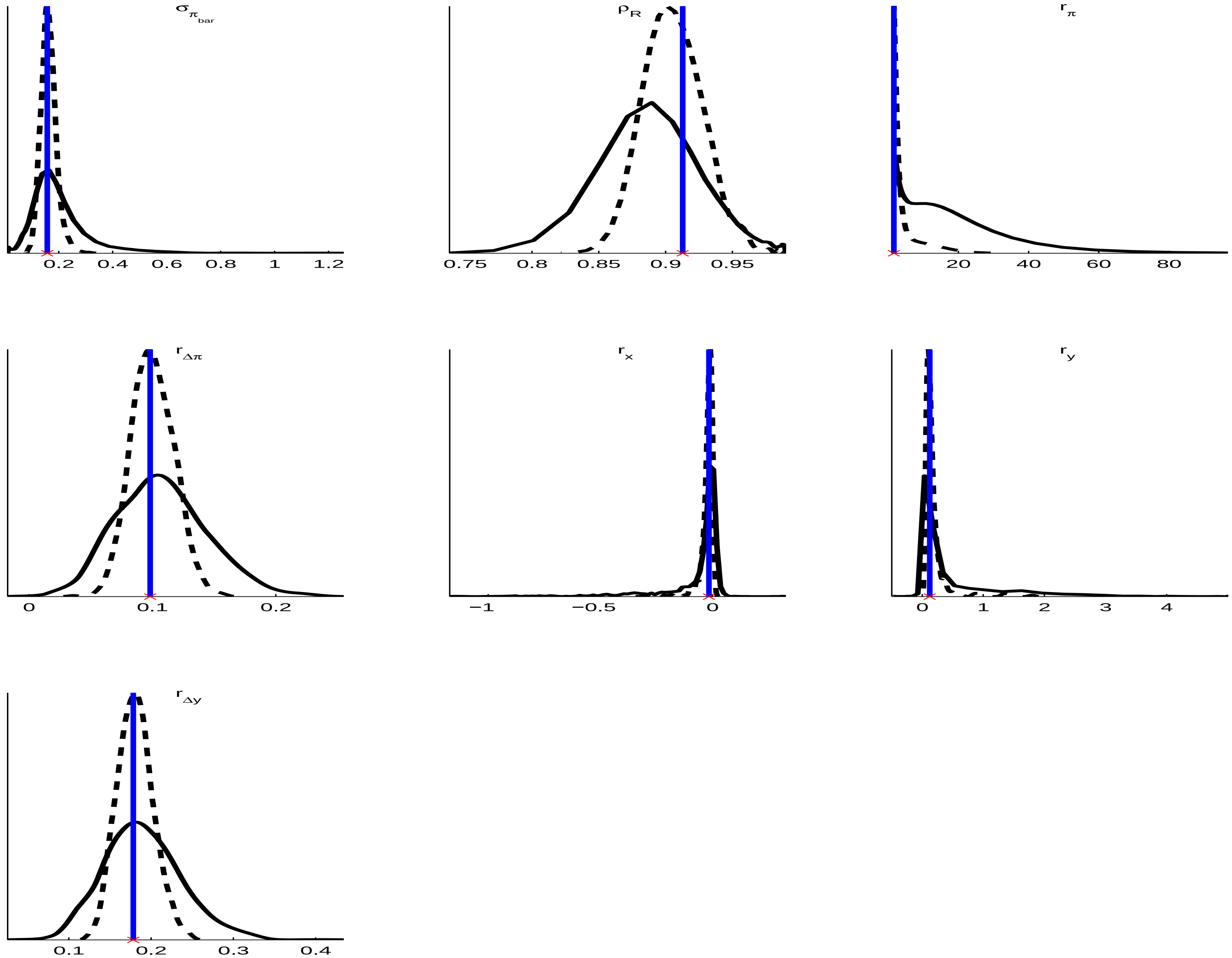

Figure 1c: Kernel density estimates of the small sample distribution for the estimates of the monetary policy parameters. The solid line shows the parameter distribution for $N = 100$, and the dashed line show the distributions for $N = 400$ observations. The vertical bar shows true parameter value, and the cross on the x-axis denotes the starting value in the optimizations.

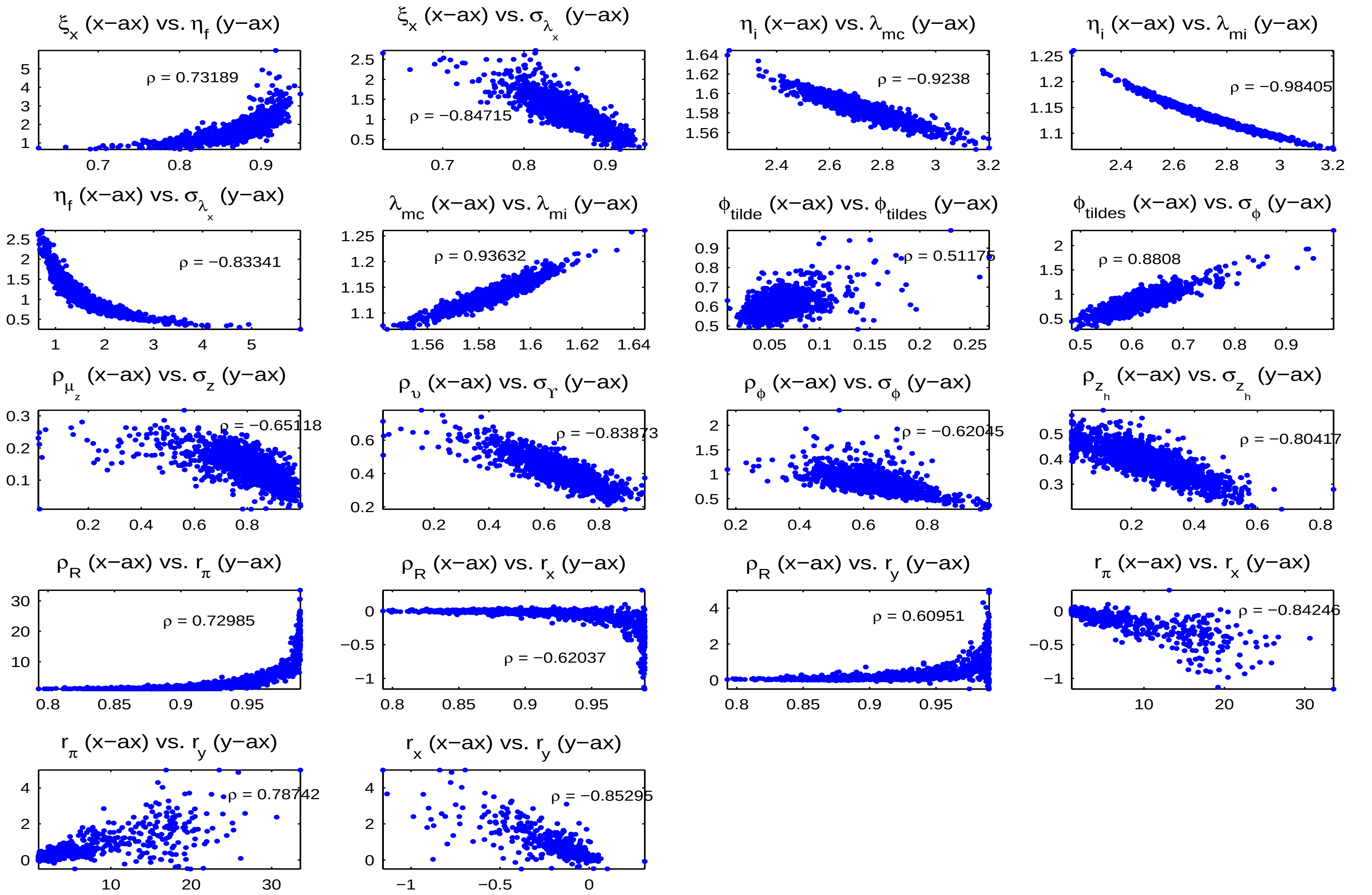

Figure 2: Pairwise estimates for parameters with cross-correlations above 0.5. $N = 100$, initializing the optimizations with the true parameters.

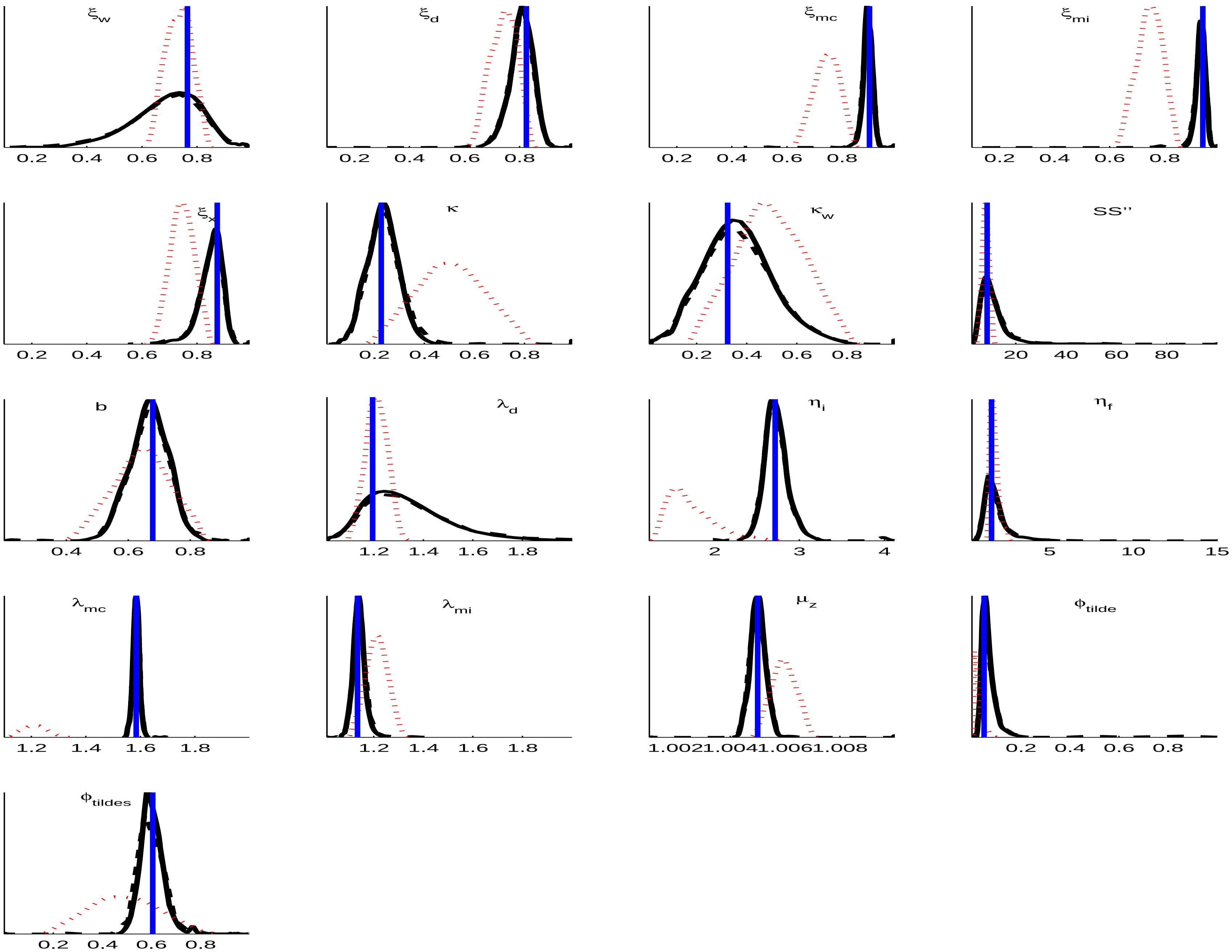

Figure 3a: Kernel density estimates of the small sample distribution for the estimates of the deep model parameters. The solid line shows the parameter distribution when initializing the estimations with the true parameters (vertical bars), and the dashed lines show the distribution when initializing the estimations using a sample from the prior (dotted).

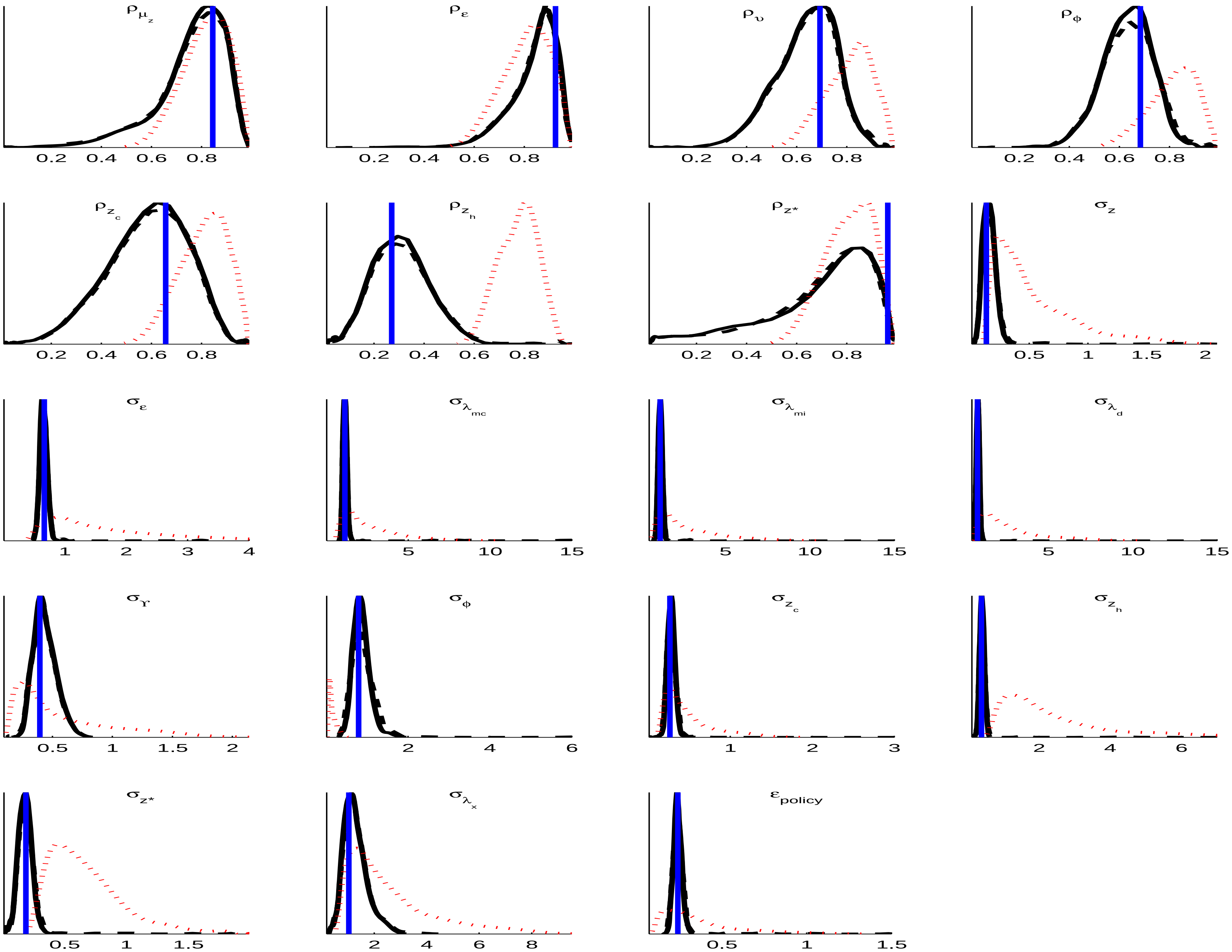

Figure 3b: Kernel density estimates of the small sample distribution for the estimates of the shock process parameters. The solid line shows the parameter distribution when initializing the estimations with the true parameters (vertical bars), and the dashed lines show the distribution when initializing the estimations using a sample from the prior (dotted).

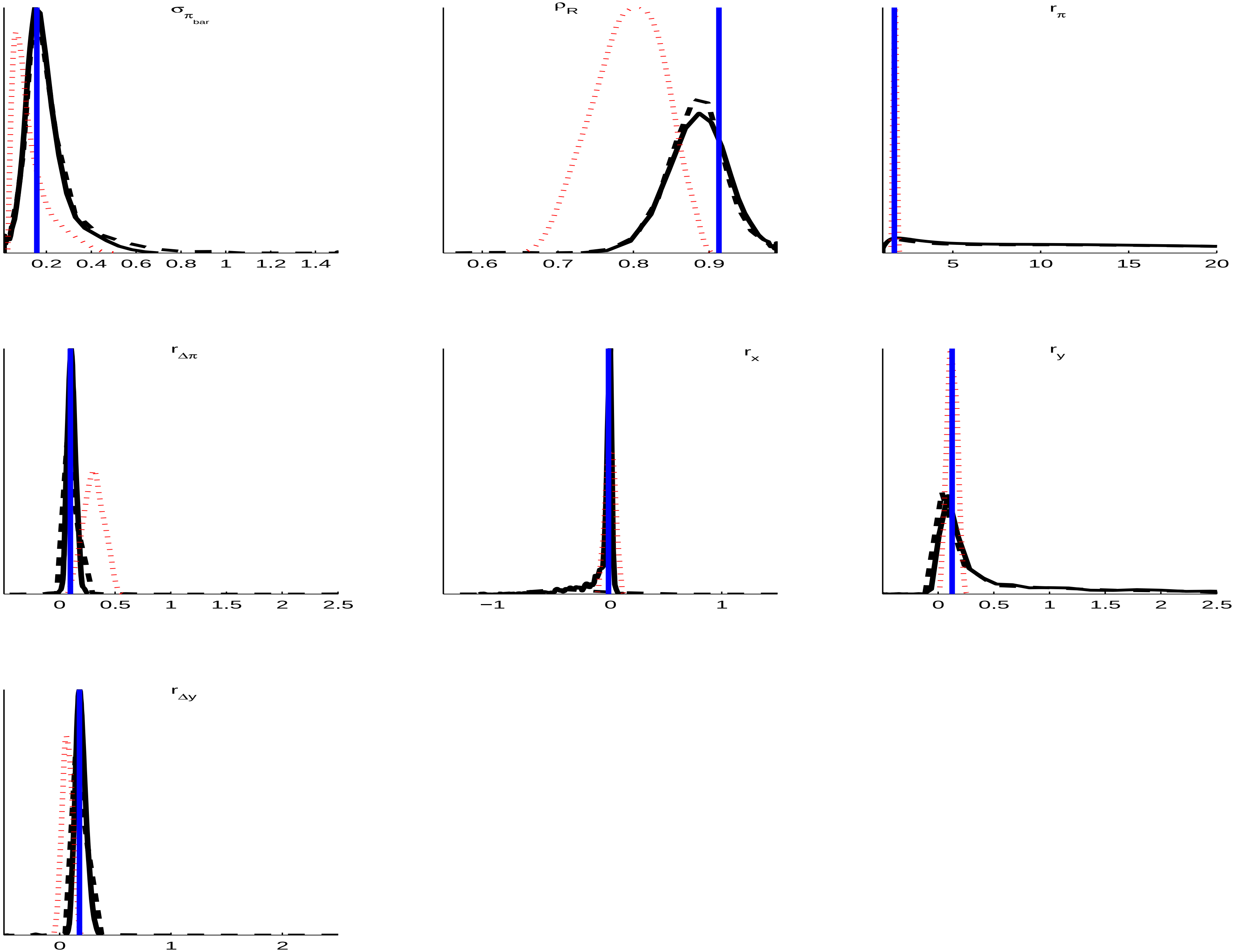

Figure 3c: Kernel density estimates of the small sample distribution for the estimates of the policy rule parameters. The solid line shows the parameter distribution when initializing the estimations with the true parameters (vertical bars), and the dashed lines show the distribution when initializing the estimations using a sample from the prior (dotted).

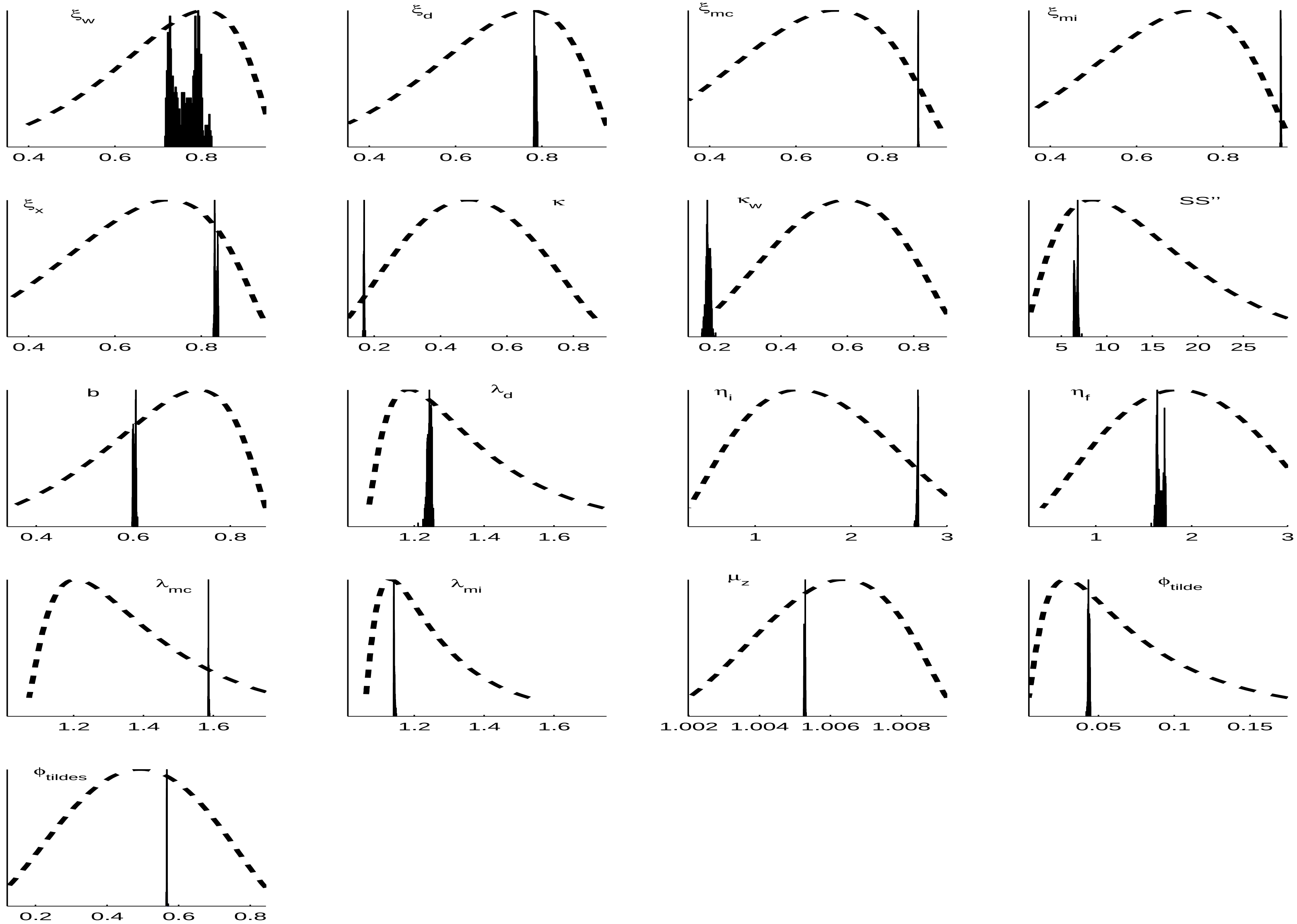

Figure 4a: Histogram for the deep model parameters estimated on one given artificial sample, based on the estimations that has converged to the same likelihood. The dashed line shows the distribution for the starting values used in the estimations.

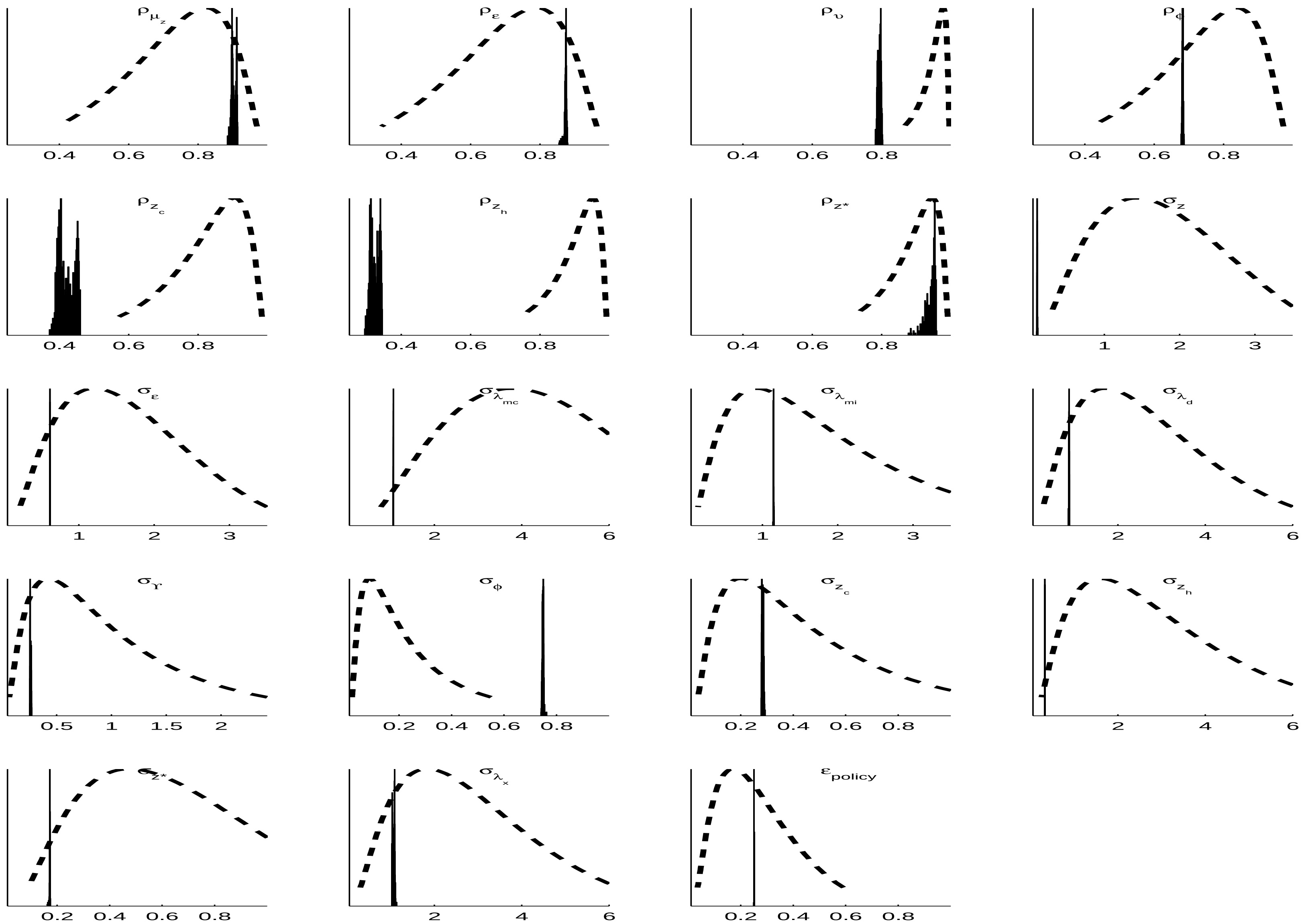

Figure 4b: Histogram for the shock process parameters estimated on one given artificial sample, based on the estimations that has converged to the same likelihood. The dashed line shows the distribution for the starting values used in the estimations.

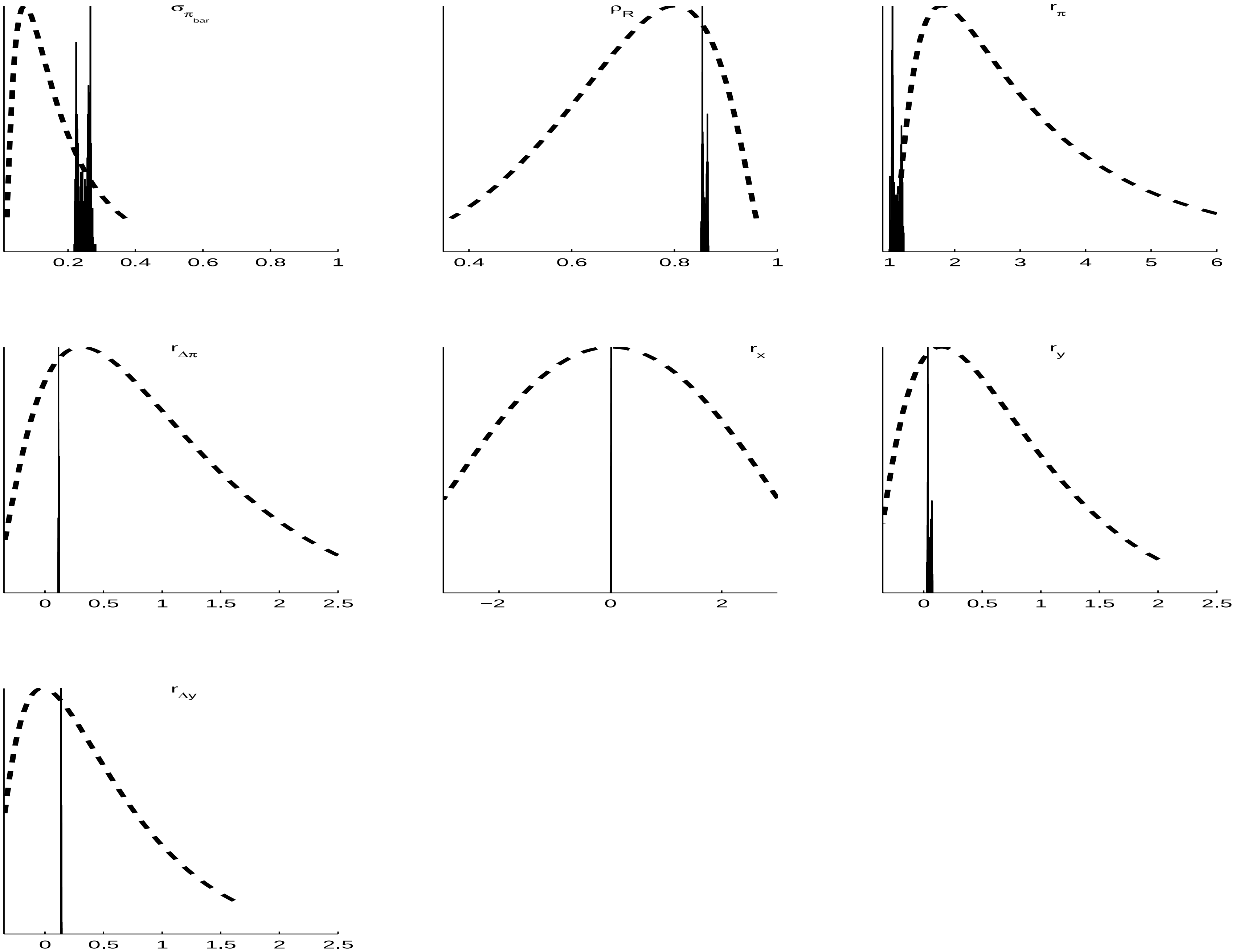

Figure 4c: Histogram for the policy rule parameters estimated on one given artificial sample, based on the estimations that has converged to the same likelihood. The dashed line shows the distribution for the starting values used in the estimations.

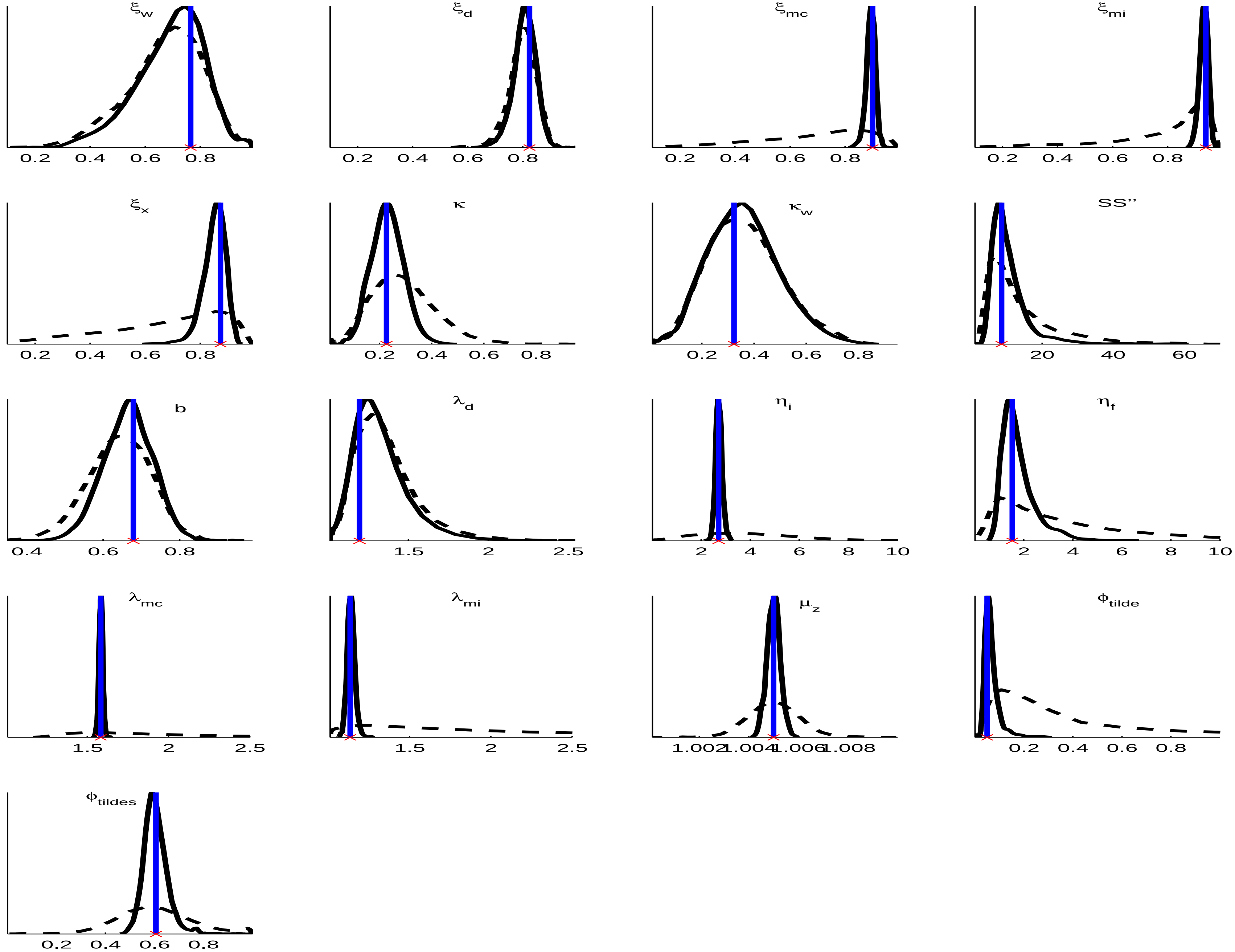

Figure 5a: Kernel density estimates of the small sample distribution for the estimates of the deep model parameters. The solid line shows the parameter distributions when the estimations are based on the full set of observable variables, and the dashed line when the estimations are based on fitting only a subset of variables (i.e., 7 "closed economy" variables). The true parameters are given by the vertical bars, and red crosses on the x-axes show starting values in the estimation.

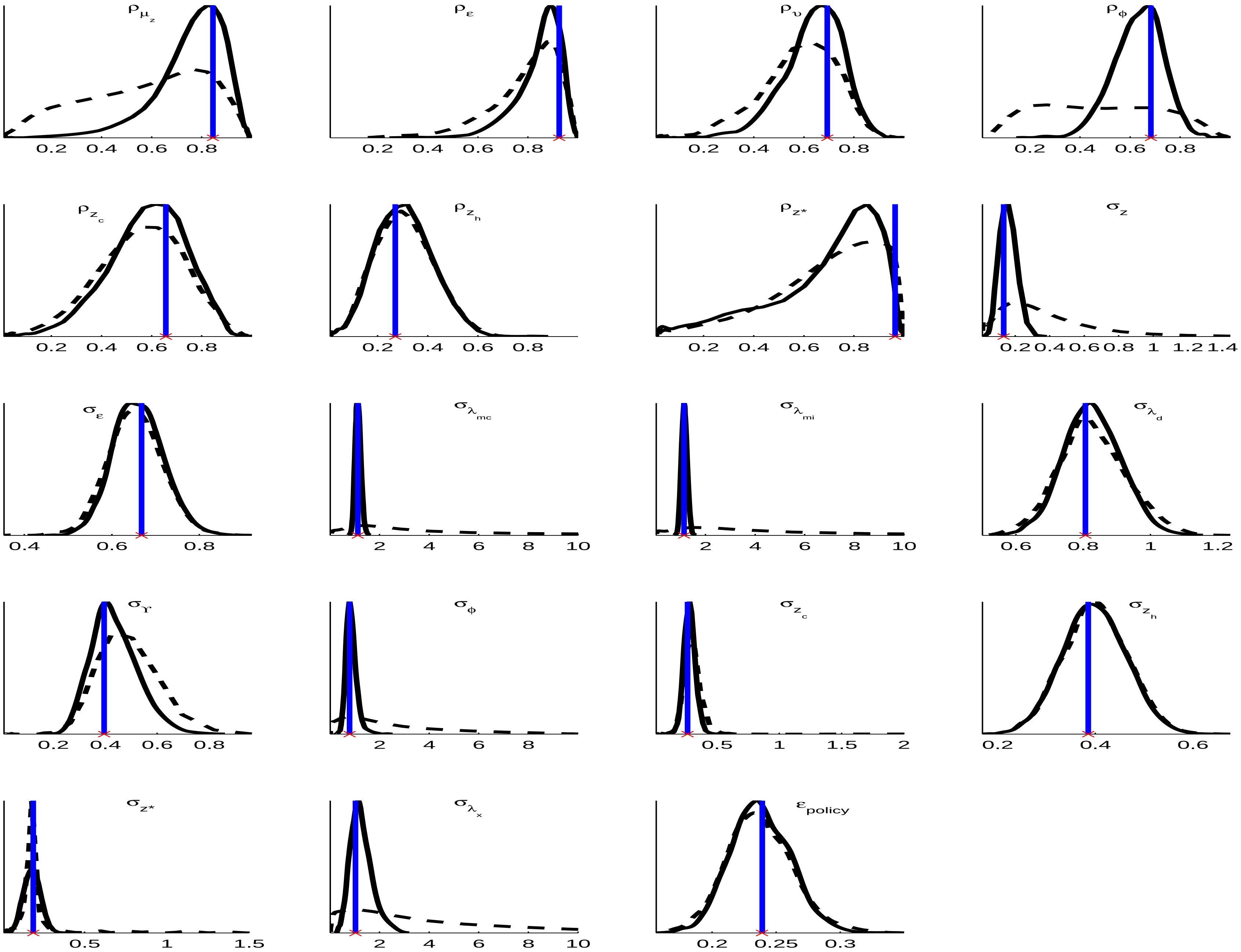

Figure 5b: Kernel density estimates of the small sample distribution for the estimates of the shock parameters. The solid line shows the parameter distribution when the estimations are based on the full set of observable variables, and the dashed line when the estimations are based on fitting only a subset of variables (i.e., 7 "closed economy" variables). The true parameters are given by the vertical bars, and red crosses on the x-axes show starting values in the estimation.

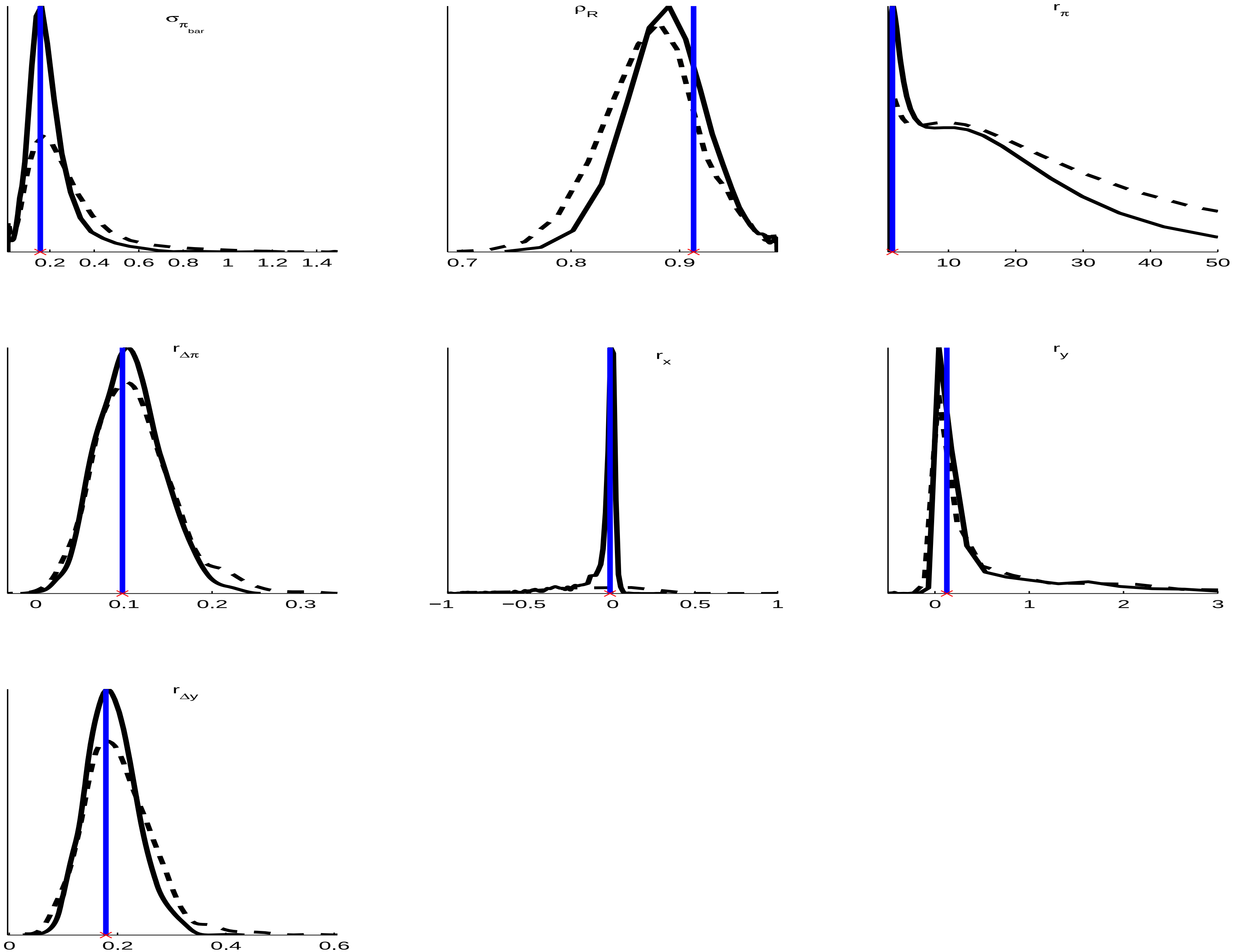

Figure 5c: Kernel density estimates of the small sample distribution for the estimates of the policy rule parameters. The solid line shows the parameter distribution when the estimations are based on the full set of observable variables, and the dashed line when the estimations are based on fitting only a subset of variables (i.e., 7 "closed economy" variables). The true parameters are given by the vertical bars, and red crosses on the x-axes show starting values in the estimation.

Appendix A. Additional simulation results

In this appendix, we present additional simulation results for three experiments.

A.1. Adding measurement errors and reestimating the fiscal and foreign VARs

In the first case, we add measurement errors to the simulated data as described in Section 3. The measurement errors are assumed to be i.i.d. and normally distributed and in the estimations they are calibrated at their true values (see Footnote 9). In addition, we also reestimate the fiscal and foreign VAR models for each sample in the same way that they are estimated on actual data rather than fixing the VAR coefficients at their true values in each simulation.

A priori, we expect this strategy, which exactly mimics the estimation strategy on actual data, to generate more dispersed parameter distributions, as the added measurement errors and estimated VARs induce additional uncertainty in the estimations. This prior is confirmed by the simulation results reported in Figures A1.a-c, where we see that the resulting parameter distributions are somewhat wider for some of the parameters. However, the key results are unaffected, the ML estimator is still unbiased in most cases and it is consistent as well.

A.2. Exploiting the co-integrating vectors in the simulations

One possible explanation to the problems with weak identification of the degree of nominal wage stickiness is that we do not exploit the cointegrating vectors when we match the model to the data. Instead of matching the variables in equation (19) where all quantities and the real wage are in quarterly growth rates, we therefore match the following set of variables instead

$$\tilde{Y}_t = \begin{matrix} [& \pi_t^d & \ln(W_t/P_t) - \ln Y_t & \ln C_t - \ln Y_t & \ln I_t - \ln Y_t & \hat{x}_t & R_t & \hat{H}_t & \Delta \ln Y_t ... \\ & & \ln \tilde{X}_t - \ln Y_t & \ln \tilde{M}_t - \ln Y_t & \pi_t^{cpi} & \pi_t^{def,i} & \ln Y_t^* - \ln Y_t & \pi_t^* & R_t^* \ ]'. \end{matrix} \tag{A.1}$$

The set of variables in (A.1) impose the model's true cointegrating vectors in the estimations, and by doing so it should provide more efficient estimation of the underlying parameters in the model.

However, as is clear from Figures A2.a-c, the efficiency gains from matching the co-integrating vectors are not very large. In most cases the resulting parameter distributions are essentially identical.

A.3. Consistency properties of ML estimator for $T = 1600$ and $T = 6400$

In Table A.1, we check the consistency properties of the ML estimator by increasing the sample size in each of the N samples to $T = 1600$ and $T = 6400$ observations. We report results for the case when we match all 15 variables in eq. (19) and the "closed economy" variables in eq. (20). As this is a very time-consuming exercise, we only report results for $N = 20$ samples for $T = 1600$ observations, and $N = 10$ samples for $T = 6400$ observations. To be able to study whether the ML estimator actually is consistent for the "closed economy" variables set, the intial values in the optimizations when we match the closed economy variables are initiated by the prior mode values in Table 2.[19]

[19] The exceptions are the persistence and standard devation for the labor supply shock process, which are intiated by their true values. The reason being that the priors for these parameters are so far off the posterior mode (i.e. the true values used to generate the samples), that they induce an implausible amount of volatility in the real wage and causes problems for the CSMINWEL algorithm.

From Table A.1, we see that the parameter distributions are collapsing at the true parameter values as $T = 6400$, but the standard deviations of the distributions indicate that the rate of convergence is substantially slower for many parameters when only the closed economy variables are matched in the estimations. A perhaps striking result is, however, that the ML estimator is actually consistent also for a relatively small set of variables. But clearly, it is much more efficient to match a larger set of variables.

Table A1: Distribution results from one particular sample of different sizes, matching two sets of variables

Parameter		True estimates	1600 observations				6400 observations			
			All variables		Closed variables		All variables		Closed variables	
			Mean of distribution	Std. of distribution	Mean of distribution	Std. of distribution	Mean of distribution	Std. of distribution	Mean of distribution	Std. of distribution
Calvo wages	ξ_w	0.77	0.74	0.07	0.76	0.02	0.77	0.01	0.77	0.01
Calvo domestic prices	ξ_d	0.83	0.82	0.02	0.83	0.01	0.83	0.00	0.83	0.00
Calvo import cons. prices	$\xi_{m,c}$	0.90	0.90	0.00	0.89	0.03	0.90	0.00	0.90	0.01
Calvo import inv. prices	$\xi_{m,i}$	0.94	0.94	0.00	0.93	0.01	0.94	0.00	0.94	0.00
Calvo export prices	ξ_x	0.87	0.87	0.01	0.85	0.04	0.87	0.00	0.87	0.02
Indexation prices	κ	0.23	0.24	0.03	0.23	0.02	0.23	0.01	0.24	0.01
Indexation wages	κ_w	0.32	0.34	0.07	0.33	0.03	0.32	0.02	0.32	0.02
Investment adj. cost	$\tilde{S}''$	8.58	8.93	0.70	8.69	0.92	8.52	0.31	8.64	0.66
Habit formation	b	0.68	0.67	0.02	0.68	0.02	0.68	0.01	0.68	0.01
Markup domestic	λ_d	1.20	1.18	0.04	1.20	0.03	1.20	0.02	1.20	0.02
Subst. elasticity invest.	η_i	2.72	2.73	0.03	2.46	0.65	2.72	0.01	2.58	0.38
Subst. elasticity foreign	η_f	1.53	1.51	0.05	1.49	0.54	1.53	0.01	1.50	0.29
Markup imported cons.	$\lambda_{m,c}$	1.58	1.58	0.00	1.56	0.07	1.58	0.00	1.58	0.05
Markup.imported invest.	$\lambda_{m,i}$	1.13	1.13	0.01	1.19	0.10	1.13	0.00	1.15	0.04
Technology growth	μ_z	1.01	1.01	0.00	1.01	0.00	1.01	0.00	1.01	0.00
Risk premium	$\tilde{\phi}$	0.05	0.05	0.00	0.06	0.01	0.05	0.00	0.05	0.01
UIP modification	$\tilde{\phi}_s$	0.61	0.58	0.03	0.63	0.04	0.61	0.00	0.61	0.02
Unit root tech. persistance	ρ_{μ_z}	0.85	0.82	0.03	0.85	0.05	0.85	0.01	0.84	0.02
Stationary tech. persistance	ρ_ε	0.93	0.92	0.01	0.92	0.01	0.92	0.01	0.92	0.01
Invest. spec. tech. persist.	ρ_Υ	0.69	0.69	0.04	0.69	0.03	0.70	0.01	0.70	0.01
Risk premium persistence	$\rho_{\tilde{\phi}}$	0.68	0.75	0.09	0.53	0.36	0.68	0.01	0.66	0.14
Consumption pref. persist.	ρ_{ζ_c}	0.66	0.73	0.06	0.65	0.04	0.65	0.01	0.64	0.02
Labour supply persistance	ρ_{ζ_h}	0.27	0.26	0.08	0.27	0.02	0.27	0.01	0.27	0.01
Asymmetric tech. persist.	$\rho_{\tilde{z}*}$	0.96	0.97	0.03	0.81	0.11	0.96	0.01	0.86	0.02
Unit root tech. shock	σ_{μ_z}	0.13	0.13	0.01	0.13	0.03	0.13	0.00	0.13	0.01
Stationary tech. shock	σ_ε	0.67	0.67	0.01	0.67	0.01	0.67	0.01	0.67	0.01
Imp. cons. markup shock	$\sigma_{\lambda_{m,c}}$	1.13	1.11	0.03	1.08	0.22	1.12	0.01	1.14	0.20
Imp. invest. markup shock	$\sigma_{\lambda_{m,i}}$	1.13	1.12	0.03	1.18	0.29	1.13	0.01	1.12	0.13
Domestic markup shock	σ_{λ_d}	0.81	0.80	0.01	0.80	0.01	0.80	0.01	0.80	0.01
Invest. spec. tech. shock	σ_Υ	0.40	0.39	0.03	0.39	0.02	0.39	0.01	0.39	0.01
Risk premium shock	$\sigma_{\tilde{\phi}}$	0.79	0.68	0.15	1.54	1.40	0.80	0.01	0.88	0.44
Consumption pref. shock	σ_{ζ_c}	0.26	0.25	0.01	0.26	0.01	0.26	0.00	0.26	0.01
Labour supply shock	σ_{ζ_h}	0.39	0.40	0.03	0.39	0.01	0.39	0.00	0.39	0.00
Asymmetric tech. shock	$\sigma_{\tilde{z}*}$	0.19	0.19	0.01	0.46	0.41	0.19	0.00	0.41	0.21
Export markup shock	σ_{λ_x}	1.03	1.05	0.03	1.02	0.51	1.02	0.01	1.06	0.22
Monetary policy shock	σ_R	0.24	0.24	0.01	0.24	0.00	0.24	0.00	0.24	0.00
Inflation target shock	$\sigma_{\bar{\pi}^c}$	0.16	0.29	0.22	0.17	0.03	0.15	0.01	0.15	0.01
Interest rate smoothing	ρ_R	0.91	0.89	0.03	0.91	0.01	0.91	0.01	0.92	0.01
Inflation response	r_π	1.67	1.31	0.44	1.56	0.31	1.65	0.14	1.73	0.18
Diff. infl response	$r_{\Delta\pi}$	0.10	0.09	0.01	0.09	0.01	0.10	0.00	0.10	0.00
Real exch. rate response	r_x	-0.02	-0.01	0.01	-0.01	0.01	-0.02	0.00	-0.02	0.01
Output response	r_y	0.13	0.08	0.05	0.12	0.05	0.12	0.02	0.13	0.02
Diff. output response	$r_{\Delta y}$	0.18	0.16	0.02	0.17	0.01	0.18	0.01	0.18	0.01

Note: The results above are based on 20 (convergenrt) estimations for the small sample (1600 obs.), and 11 (convergent) estimations for the large sample (6400 obs.). The optimizations are intialized by the prior mode values in Table 2 for the closed variables set and for the all variables set by the true parameters except for the shocks which are intialized by their prior mode values.

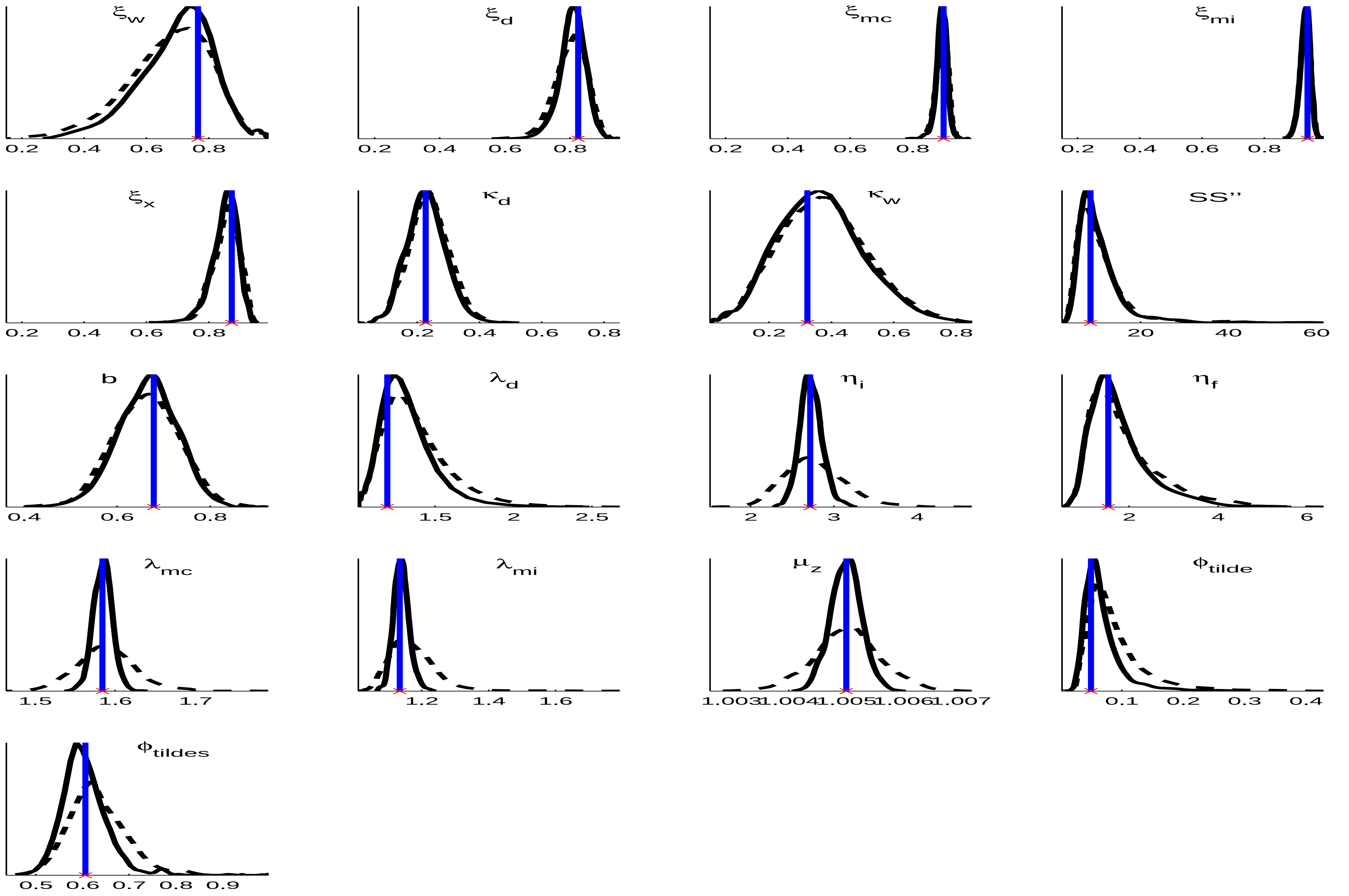

Figure A1a: Kernel density estimates of the small sample distribution for the estimates of the deep model parameters. The solid line shows the parameter distribution when estimating the model without measurement errors and keeping the foreign and fiscal VAR models fixed, and the dashed line shows the distribution when estimating the model with measurement errors and reestimating the foreign and fiscal VAR models for each sample. The vertical bar shows true parameter value, and the cross on the x-axis denotes the starting value in the optimizations. $N = 100$ observations.

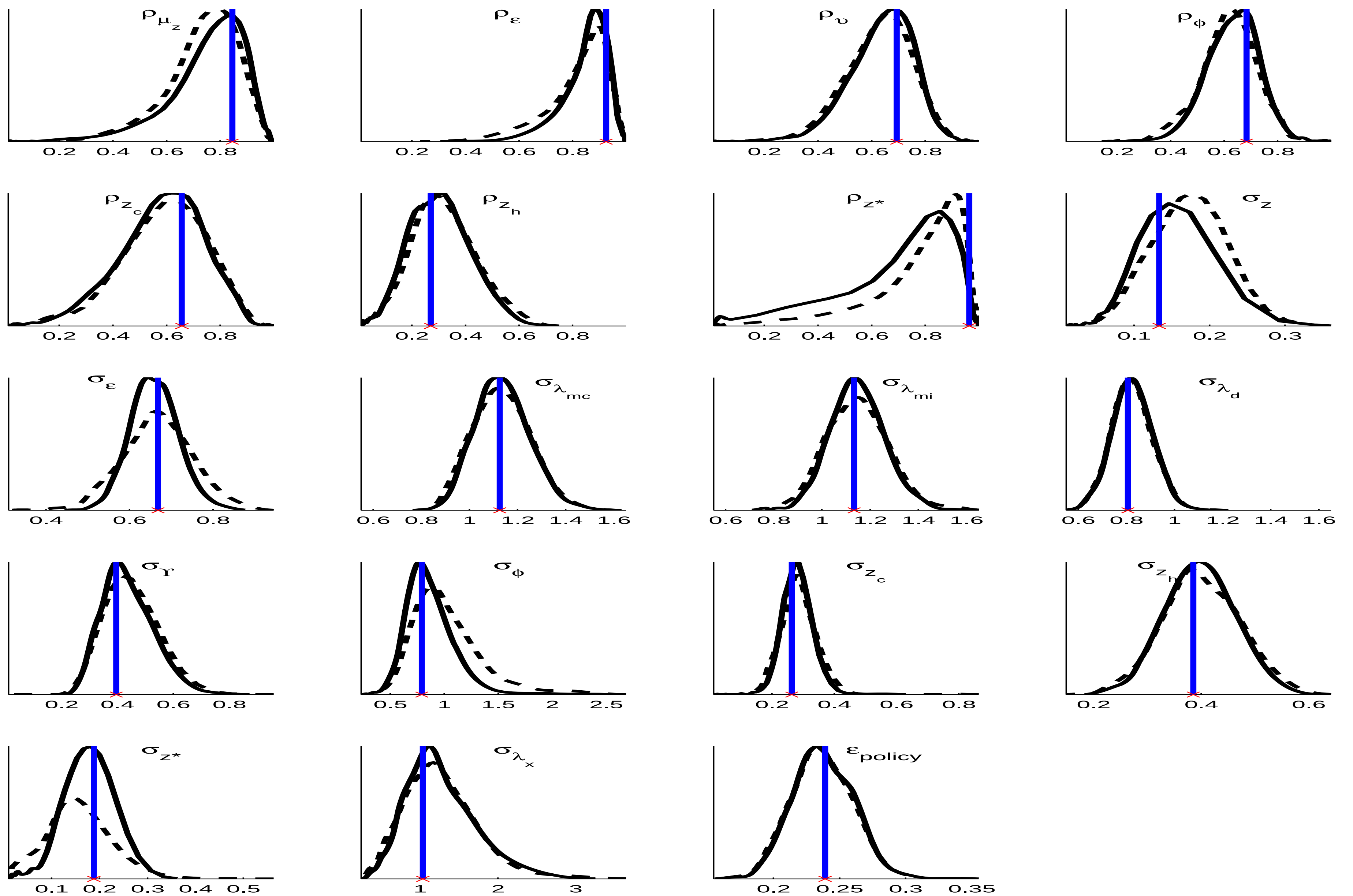

Figure A1b: Kernel density estimates of the small sample distribution for the estimates of the shock parameters. The solid line shows the parameter distribution when estimating the model without measurement errors and keeping the foreign and fiscal VAR models fixed, and the dashed line shows the distribution when estimating the model with measurement errors and reestimating the foreign and fiscal VAR models for each sample. The vertical bar shows true parameter value, and the cross on the x-axis denotes the starting value in the optimizations. $N = 100$ observations.

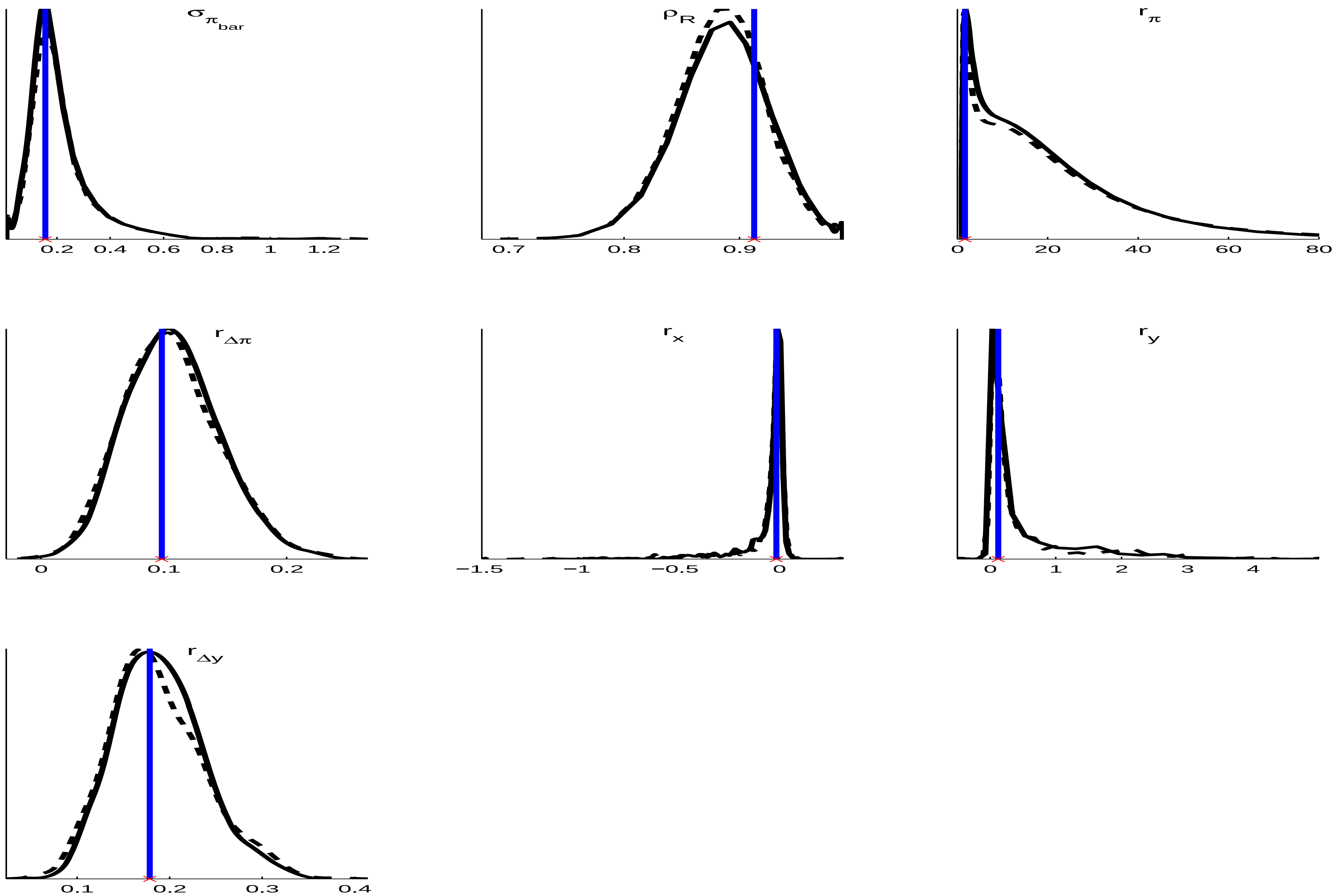

Figure A1c: Kernel density estimates of the small sample distribution for the estimates of the policy parameters. The solid line shows the parameter distribution when estimating the model without measurement errors and keeping the foreign and fiscal VAR models fixed, and the dashed line shows the distribution when estimating the model with measurement errors and reestimating the foreign and fiscal VAR models for each sample. The vertical bar shows true parameter value, and the cross on the x-axis denotes the starting value in the optimizations. $N = 100$ observations.

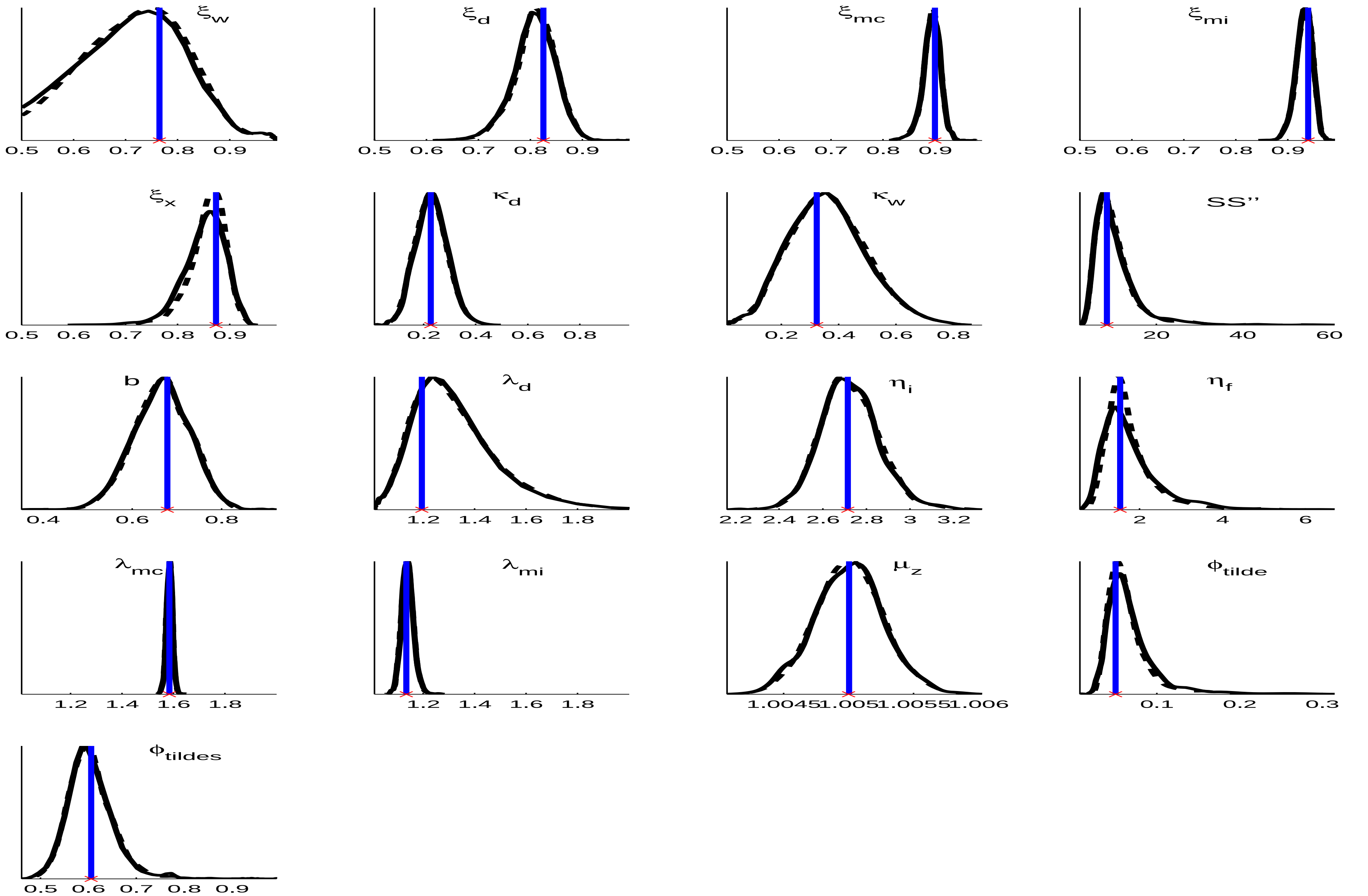

Figure A2a: Kernel density estimates of the small sample distribution for the estimates of the deep model parameters. The solid line shows the parameter distribution when estimating the model in first differences, and the dashed line shows the distribution when estimating the model using the true cointegrating vectors. The vertical bar shows true parameter value, and the cross on the x-axis denotes the starting value in the optimizations. $N = 100$ observations.

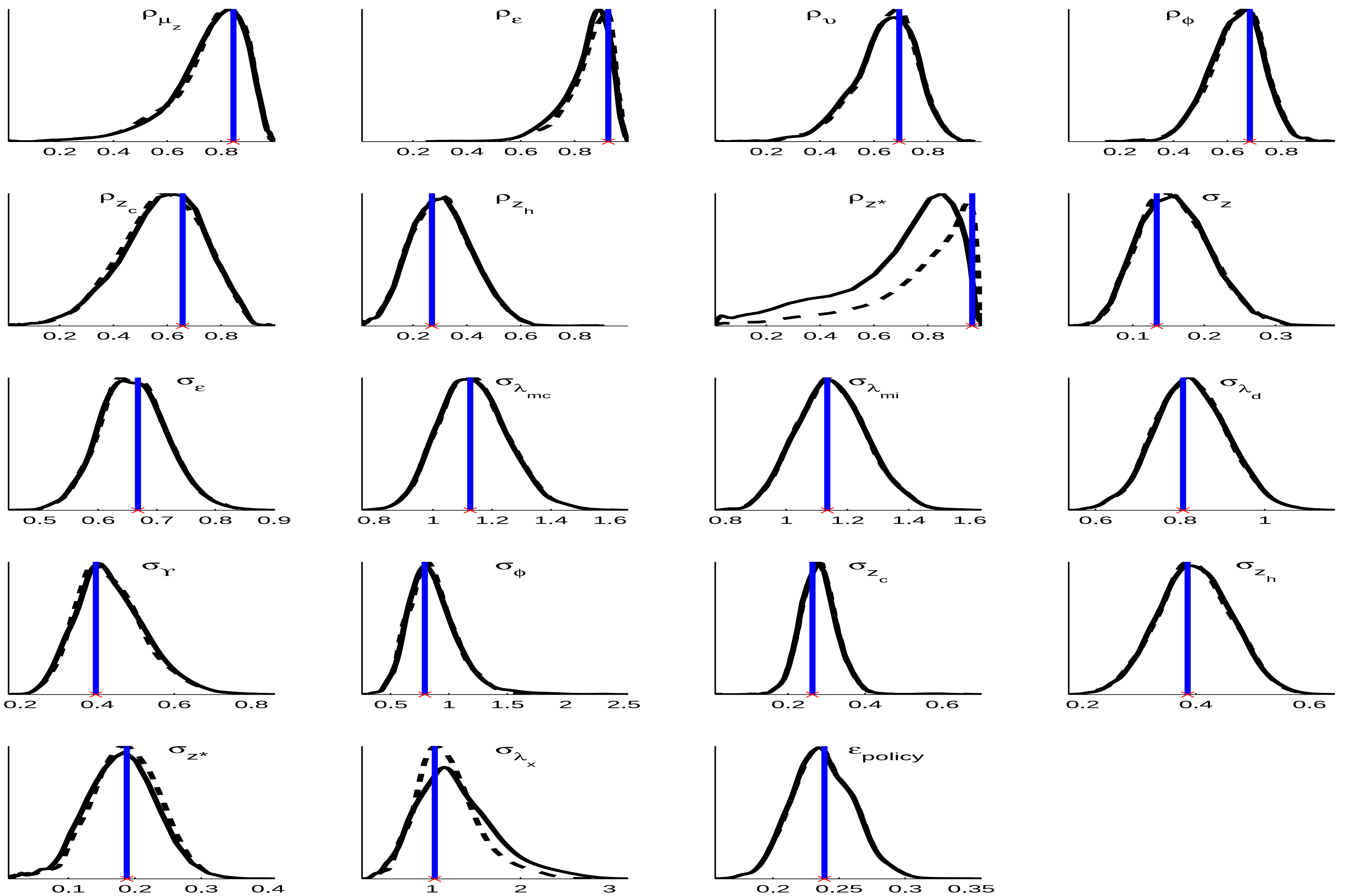

Figure A2b: Kernel density estimates of the small sample distribution for the estimates of the shock parameters. The solid line shows the parameter distribution when estimating the model in first differences, and the dashed line shows the distribution when estimating the model using the true cointegrating vectors. The vertical bar shows true parameter value, and the cross on the x-axis denotes the starting value in the optimizations. $N = 100$ observations.

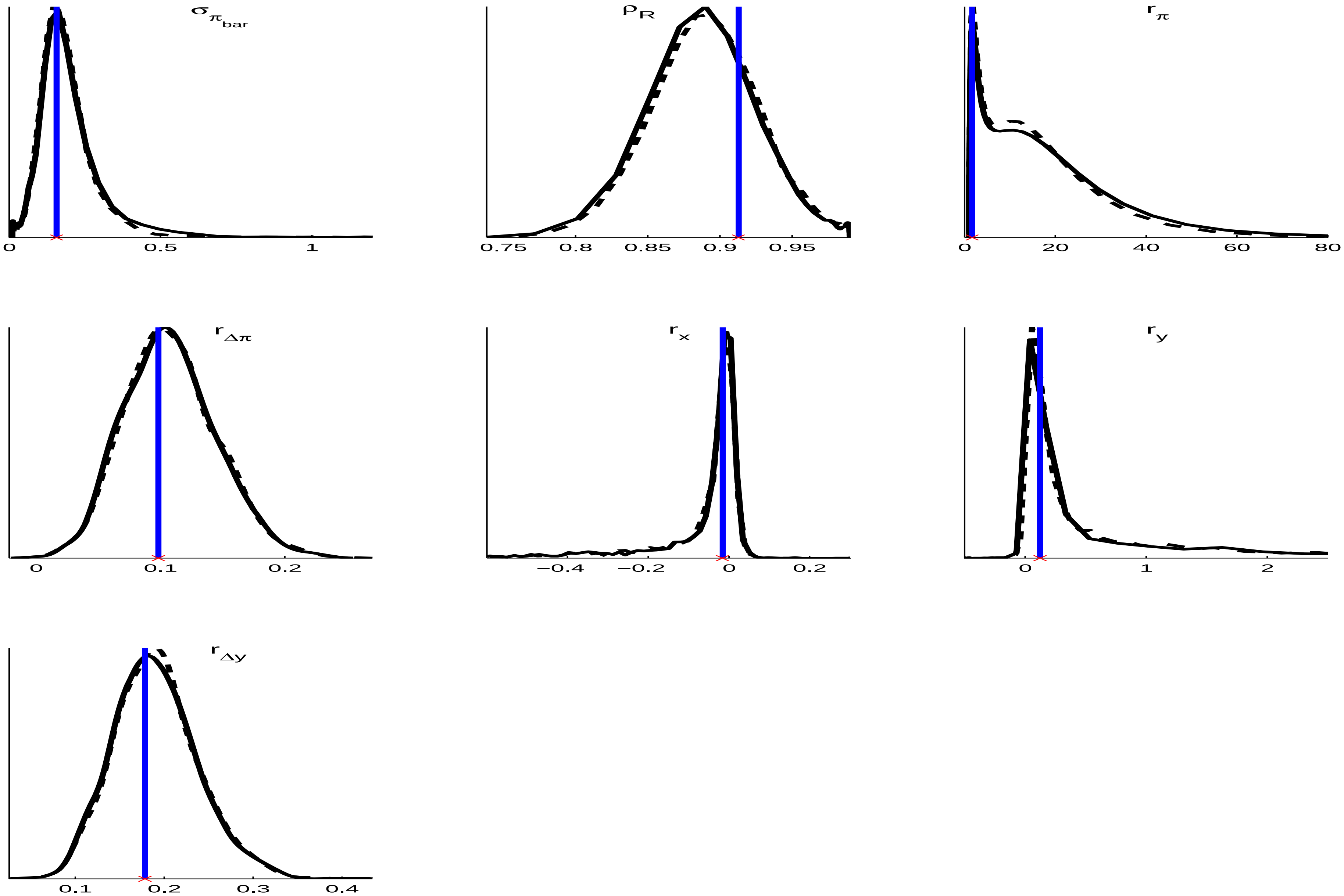

Figure A2c: Kernel density estimates of the small sample distribution for the estimates of the monetary policy parameters. The solid line shows the parameter distribution when estimating the model in first differences, and the dashed line shows the distribution when estimating the model using the true cointegrating vectors. The vertical bar shows true parameter value, and the cross on the x-axis denotes the starting value in the optimizations. $N = 100$ observations.