EFSF
7/12/07
2:00 PM

ESTIMATION OF DSGE MODELS WHEN THE DATA ARE PERSISTENT

Yuriy Gorodnichenko *   Serena Ng†

July 2, 2007

Preliminary and incomplete
Comments are welcome

Abstract

An active area of research in macroeconomics is to take DSGE models to the data. Much of the focus has been on estimation and testing of models solved under specific assumptions about how the exogenous variables grow over time. In this paper, we first show that if the trends assumed for the model are incompatible with the observed data, or that the detrended data used in estimation are inconsistent with the stationarity concepts of the model, the estimates can be severely biased even in large samples. Linearly detrending a unit root process can lead to non-standard inference as the regressors are not stationary, while the regression can be spurious if the data are inappropriately filtered. We then suggest a quasi-differencing approach that is robust to whether shocks in the model are assumed to be permanent or transitory. Root-T consistent and asymptotically normally estimates can be obtained without requiring the researcher to take a stand on the dynamic properties of the data. Simulations also show that the methodology works well when the shock process is correctly assumed and is far more accurate than standard estimators when the model parameter is near the unit circle. These properties hold even when there are multiple shocks.

*Department of Economics, University of Michigan, Ann Arbor, MI 48109 Email: ygorodni@umich.edu.
†Department of Economics, University of Michigan, Ann Arbor, MI 48109 Email: Serena.Ng@umich.edu.

1 Introduction

Dynamic stochastic general equilibrium (DSGE) models are now accepted as the primary framework for macroeconomic analysis. Until recently, counterfactual experiments were conducted by assigning the parameters of the models with values that are loosely calibrated to the data. More recently, serious efforts have been made to estimate the model parameters using classical and Bayesian methods. This permits researchers to assess how well the models fit the data both in and out of samples. Formal estimation also permits errors arising from sampling or model uncertainty to be explicitly accounted for in counterfactual policy simulations. Arguably, DSGE models are now taken more seriously as a tool for policy analysis because of such serious econometric investigations.

This paper points out two potential problems specific to the estimation of DSGE models when either the data and/or the model variables are persistent or non-stationary. When one or more forcing processes in a DSGE model are non-stationary, the model variables in level form have to be first normalized by appropriate trending variables. The variables in the log-linearized model are then interpreted as deviations from the steady state of the normalized variables. In order to take the model to the data, a researcher must construct data analogs of the model concepts, and in doing so, must choose a method for detrending the data. The first problem arises when the method of detrending does not agree with the definition of the trends in the model. The second problem arises when the data are detrended to match the model concepts but that the empirically detrended data remain non-stationary or are over-differenced. Both issues can pose problems for estimation and inference. Hereafter, we refer to these issues as Data Detrending (DD) and Model Trend Specification (MTS) problems. A concise overview of the issues associated with estimating DSGE models is as follows:

	Step 1		Step 2		Step 3
	Model Specification	$\rightarrow$	Data Detrending	$\rightarrow$	Estimation
Problems:		MTS		DD	

Problem (DD) is concerned with how the observed data are filtered. The filtered data can be stationary and yet the trends associated with the stationary component of the data can be inconsistent with how the trends are defined in the model. Problem (DD) would most likely arise when a researcher detrends the data to ensure that the deviation from a trend component is stationary, but is unaware that the trends that accomplish this task are inconsistent with the trends specified in the model. For example, the model may specify the trend as a random walk, but the data may be detrended by a two-sided symmetric filter. Whereas the stationary component in the model is white noise, the filtered series can be serially correlated. In this case, the error term associated with the empirical Euler equations can be serially correlated. The moment conditions used to estimate the parameters will not be zero even in the population.

Problem (MTS) is concerned with whether the assumption about the trend in the model is consistent with the trend in the data. Problem (MTS) is often related to whether the detrended data are stationary. This issue can arise if, for example, the model assumes that technology is trend stationary and thus the data are linearly detrended accordingly. However, the detrended data will still be non-stationary if in fact the data contain stochastic trends. As is well known, classical inference procedures can be misleading when the regressors are non-stationary or highly persistent, and estimation of a spurious regression cannot be ruled out. An additional issue that confronts researchers is that in finite samples, it is very difficult to ascertain whether the data are stationary or not. Yet, existing estimators of DSGE models require that the researcher takes a stand on the stationarity property of the data.

The two problems are not unrelated. For instance, a mistake in the first step (i.e., MTS problem) can distort filtering and estimation results even if there is no Problem (DD), i.e., the researcher uses a model consistent trend to detrend the data. In any case, an error in either model trend specification or data detrending can seriously distort the results in the estimation step.

Table 1 is a non-exhaustive listing of how trends are treated in some notable papers. While there are exceptions, the majority of the analysis assumes that non-stationarity in the models is due to a deterministic trend. The empirical analysis then proceeds to estimate the models on linearly detrended data. However, as Nelson and Plosser (1982) pointed out, the assumption of trend stationarity for variables such as real output is questionable. Stochastic trends are assumed in some studies and the first differenced data are then used in estimation. While much is known about estimation and inference of linear models with non-stationary data, little is known about how the treatment of trends affects estimation of DSGE models. This paper sheds some light on this issue.

This paper is intimately related to previous literature investigating properties of filtered data. From this literature we know that improper filtering can alter persistence and volatility of the series (e.g., Cogley and Nason (1995)), induce spurious correlations in the filtered data (e.g., Harvey and Jaeger (n.d.)), change error structure (e.g., Singleton (1988)), distort inference (e.g., Christiano and den Haan (1996)) or even yield non-stationary series (e.g., Nelson and Kang (1981)). However, much of this literature is focused on univariate analysis and relatively little is known about effects of filtering on estimates of structural parameters in DSGE models. The systems approach provides a complete characterization of the model and thus the estimates are more efficient if the model is correctly specified. But mis-specification in one equation can affect estimates in other equations. In an early contribution, King and Rebelo (1993) simulate an RBC model and show that HP filtered data are qualitatively different from the raw data. Although these authors do not estimate the model on filtered data, they hint that estimates of structural parameters can be adversely affected by filtering. From our non-exhaustive review of the literature, it appears that analysis of Problem (DD) has been predominantly constrained to univariate framework or informal discussions about possible effects in estimating DSGE models.

In a similar vein, analysis of Problem (MTS) has been generally conducted within the univariate framework. In a study closely related to ours, Cogley (2001) investigates formally how Problem (MTS) can affect the estimates of structural parameters. He shows that inappropriate choice of trend (i.e., trend stationary versus difference stationarity forcing variables) can lead to strong biases in the maximum likelihood estimates. He considers several possibilities to circumvent Problem (MTS) and finds that using cointegration relationships in unconditional Euler equations works the best since in this formulation moments used in GMM estimation remain stationary irrespective of whether the data are trend or difference stationary. Our approach, which is based on estimating covariance structures, is different from and complementary to Cogley's approach.[1] Instead of comparing estimators, we study the properties of the covariance structure estimator alone to focus on the sensitivity of the estimates to the model underlying the covariance structure as well as the to the choice of sample analogues of the model variables.

Specifically, we propose a robust strategy to handle uncertainty as to whether the data are trend or difference stationary. Under the proposed approach, the model is solved in a way that does not require the researcher to take a stand on whether the shocks have permanent or transitory effects on the model variables. The key is to solve the model in quasi-differenced form *and* proceed to estimation using quasi-differenced data. The quasi-differencing approach is shown to be effective even when there are multiple shocks, of which a subset of them may be permanent. Although our analysis is motivated as classical estimator, it can be adapted into a Bayesian framework.

We use a stochastic growth model to illustrate the problems under consideration. When the trends assumed for the model agree with the trends present in the data, the estimated parameters are mean and median unbiased. Otherwise, the estimates can deviate significantly from the true values. For instance, even though the HP filter can remove both linear and stochastic trends, the detrended data no longer have the same dynamics as those implied by the model. As a result,

[1] Combination of these two approaches is straightforward and, although we do not investigate the benefits on combining cointegration and quasi-differencing formally, one may expect that such a combination can be quite fruitful in sharpening the estimates.

the HP estimates can be severely biased. Estimates of parameters governing the propagation and amplification mechanisms in the model can be greatly distorted or poorly identified.

The structure of the paper is as follows. In the next section, we lay out a standard neoclassical growth model. We linearize the model and show how one can solve it under different assumptions about trends in the forcing variables. We present the estimation procedure and illustrate Problems (DD) and (MTS) with a few specific examples. In Section 3, we report simulation results. In particular, we demonstrate a superior performance of our robust approach and distortions in popular estimators due to Problems (DD) and (MTS). We consider several extensions of the baseline growth model to highlight the issues associated with estimation of endogenous propagation/amplification mechanisms and estimation of models with multiple structural shocks. We also briefly contrast statistical properties of our approach and popular alternatives. In Section 4, we develop a general framework for using quasi-differencing to estimate structural parameters in linearized DSGE models. We conclude in Section 5.

2 An Example: Neoclassical Growth Model

2.1 The General Setup

We use the stochastic growth model to highlight the problems under investigation. The problem facing the central planner is:

$$\max E_t \sum_{t=0}^{\infty} \beta^t \Big(\ln C_t - \theta(L_t/Q_t) \Big)$$

subject to

$$\begin{aligned}
Y_t &= C_t + I_t = K_{t-1}^{\alpha}(Z_t L_t)^{(1-\alpha)} \\
K_t &= (1-\delta)K_{t-1} + I_t \\
Z_t &= \exp(\bar{g}t)\exp(u_t^z), \quad u_t^z = \rho_z u_{t-1}^z + e_t^z, \quad |\rho_z| \leq 1 \\
Q_t &= \exp(u_t^q), \qquad\qquad u_t^q = \rho_q u_{t-1}^q + e_t^q, \quad |\rho_q| < 1.
\end{aligned}$$

where Y_t is output, C_t is consumption, K_t is capital, L_t is labor input, Z_t is the level of technology, and Q_t is a labor supply shock. In the general case, ρ_q and ρ_z can be on the unit circle. The first order conditions are:

$$\begin{aligned}
\theta C_t &= (1-\alpha)K_{t-1}^{\alpha} Z_t^{(1-\alpha)} L_t^{-\alpha} Q_t \\
E_t C_{t+1} &= \beta C_t \Big(\alpha K_t^{\alpha-1}(Z_{t+1}L_{t+1})^{(1-\alpha)} + (1-\delta) \Big) \\
K_{t-1}^{\alpha}(Z_t L_t)^{(1-\alpha)} &= C_t + K_t - (1-\delta)K_{t-1}
\end{aligned}$$

If $\bar{g} = 0$ and $|\rho_z|, |\rho_q| < 1$, then under regularity conditions, a solution for model log-linearized around the steady state values exists. But once technology is allowed to grow over time, the model solution as well as the estimation approach depends on the properties of Z_t and Q_t.

Let lower case letters denote the natural logarithm of the variables, e.g. $c_t = \log C_t$. Let c_t^*, be such that in $c_t - c_t^*$ is stationary; k_t^* and z_t^* are similarly defined. Note that c_t^* and k_t^* are model concepts. Hereafter, we will use DT and ST to refer to the case when $|\rho_z| < 1$ and $|\rho_z| = 1$, respectively. The assumption on $|\rho_q|$ will vary depending on the context. Where appropriate, we will drop Q_t to simplify the analysis.

2.2 Solving the One Shock Model

To fix ideas, suppose for now that technology is the only shock in the system. Hence, Q_t is suppressed. We consider separately when $|\rho_z| < 1$ and when $|\rho_z| = 1$.

When $|\rho_z| < 1$, $c_t^* = k_t^* = \bar{g}t$. The detrended variables in the model are then defined as $\widehat{c}_t = c_t - c_t^*$, $\widehat{k}_t = k_t - k_t^*$ and $\widehat{l}_t = l_t$.[2] The log-linearized model in terms of $\widehat{c}_t, \widehat{k}_t, \widehat{l}_t$ is

DT Model

$$E_t \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & A_0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \widehat{c}_{t+1} \\ \widehat{k}_{t+1} \\ \widehat{l}_{t+1} \end{bmatrix} = \begin{bmatrix} -1 & 0 & -\alpha \\ 1 & A_0 & 0 \\ A_1 & A_2 & \alpha - 1 \end{bmatrix} \begin{bmatrix} \widehat{c}_t \\ \widehat{k}_t \\ \widehat{l}_t \end{bmatrix} + \begin{bmatrix} 0 & \alpha & 0 \\ 0 & 0 & 0 \\ 0 & A_4 & 0 \end{bmatrix} \begin{bmatrix} \widehat{c}_{t-1} \\ \widehat{k}_{t-1} \\ \widehat{l}_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ -A_0 \\ 0 \end{bmatrix} u_{t+1}^z + \begin{bmatrix} 1-\alpha \\ 0 \\ \alpha - 1 \end{bmatrix} u_t^z \quad (1)$$

where we suppress constant terms and define entries of the matrices in (1) as follows

$$\begin{aligned} A_0^* &= 1 - \beta \frac{1-\delta}{1+\bar{g}} \\ A_0 &= (\alpha - 1) A_0^* \\ A_4 &= -\alpha - (1-\delta) A_3, \\ A_3 &= \frac{\alpha \beta}{(1+\bar{g}) A_0^*}, \\ A_2 &= (1+\bar{g}) A_3, \\ 1 &= A_1 + A_2 - (1-\delta) A_3. \end{aligned}$$

We will refer to (1) as the trend stationary (DT) representation of the model. Let $\widehat{m}_t = (\widehat{c}_t, \widehat{k}_t, \widehat{l}_t)'$. Since a shock to technology has temporary effects, $\widehat{m}_t$ is stationary. We can compactly write (1) as

$$E_t \Gamma_2^D \widehat{m}_{t+1} = \Gamma_0^D \widehat{m}_t + \Gamma_1^D \widehat{m}_{t-1} + \Psi_1^D u_{t+1}^z + \Psi_0^D u_t^z.$$

[2] Note that labor L_t is stationary for all $|\rho_z| \leq 1$ and thus we do not need to scale it.

The QZ decomposition or similar methods are used to solve the system of expectation equations for the reduced form.

When $|\rho_z| = 1$, the log-linearized model is expressed in $\widetilde{c}_t$, $\widetilde{k}_t$, and $\widetilde{k}_t$, where

$$\widetilde{c}_t = c_t - z_t, \quad \widetilde{k}_t = k_t - z_t, \quad \widetilde{l}_t = l_t.$$

The model is now represented by the following system of expectational equations:

ST Model

$$E_t \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & A_0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \widetilde{c}_{t+1} \\ \widetilde{k}_{t+1} \\ \widetilde{l}_{t+1} \end{bmatrix} = \begin{bmatrix} -1 & 0 & -\alpha \\ 1 & A_0 & 0 \\ A_1 & A_2 & \alpha - 1 \end{bmatrix} \begin{bmatrix} \widetilde{c}_t \\ \widetilde{k}_t \\ \widetilde{l}_t \end{bmatrix} + \begin{bmatrix} 0 & \alpha & 0 \\ 0 & 0 & 0 \\ 0 & A_4 & 0 \end{bmatrix} \begin{bmatrix} \widetilde{c}_{t-1} \\ \widetilde{k}_{t-1} \\ \widetilde{l}_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ -A_0 \\ 0 \end{bmatrix} e^z_{t+1} + \begin{bmatrix} 1-\alpha \\ 0 \\ \alpha - 1 \end{bmatrix} e^z_t \tag{2}$$

We will refer to (2) as the stochastic trend (ST) representation of the model. Let $\widetilde{m}_t = (\widetilde{c}_t, \widetilde{k}_t, \widetilde{l}_t)'$ and compactly write the system as

$$E_t \Gamma^S_2 \widetilde{m}_{t+1} = \Gamma^S_0 \widetilde{m}_t + \Gamma^S_1 \widetilde{m}_{t-1} + \Psi^S_1 e^z_{t+1} + \Psi^S_0 e^z_t.$$

Now $\widetilde{m}_t$ and $\widehat{m}_t$ are related as follows:

$$\widetilde{c}_t = \widehat{c}_t - u^z_t, \quad \widetilde{k}_t = \widehat{k}_t - u^z_t, \quad \widetilde{l}_t = \widehat{l}_t.$$

Effectively, subtracting u^z_t from appropriate variables as in the ST model changes the object of interest in the model from $\widehat{m}_t$ (which would not be stationary under ST) to $\widetilde{m}_t$ (which is stationary under ST).

An alternative to solving the ST model in terms of $\widetilde{m}_t$ is to consider the first-differenced representation of the DT model.

$\Delta^1 DT$ Model

$$E_t \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & A_0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta^1\widehat{c}_{t+1} \\ \Delta^1\widehat{k}_{t+1} \\ \Delta^1\widehat{l}_{t+1} \end{bmatrix} = \begin{bmatrix} -1 & 0 & -\alpha \\ 1 & A_0 & 0 \\ A_1 & A_2 & \alpha - 1 \end{bmatrix} \begin{bmatrix} \Delta^1\widehat{c}_t \\ \Delta^1\widehat{k}_t \\ \Delta^1\widehat{l}_t \end{bmatrix} + \begin{bmatrix} 0 & \alpha & 0 \\ 0 & 0 & 0 \\ 0 & A_4 & 0 \end{bmatrix} \begin{bmatrix} \Delta^1\widehat{c}_{t-1} \\ \Delta^1\widehat{k}_{t-1} \\ \Delta^1\widehat{l}_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ -A_0 \\ 0 \end{bmatrix} e^z_{t+1} + \begin{bmatrix} 1-\alpha \\ 0 \\ \alpha - 1 \end{bmatrix} e^z_t \tag{3}$$

Here the superscript "1" in Δ^1 emphasizes that ρ_z is constrained to be equal to one. Clearly, first differencing removes the permanent shock in $\widehat{m}_t$, while $\widetilde{m}_t$ subtracts the permanent shock from $\widehat{m}_t$. Not surprisingly, (2) and (3) both yield stationary solutions to the ST model.[3]

The system of equations (1), (2) and (3) both correspond to the same stochastic growth model. Not surprisingly, the rational expectations solution (that is, QZ decomposition) for variables in levels is the same irrespective of which model we solve. The models are distinguished only in what variables we analyze. In other words, although the model is defined in terms of $\widehat{m}_t$ for DT, $\Delta\widehat{m}_t$ for Δ^1DT, and $\widetilde{m}_t$ for ST, the solution for the unique rational expectations equilibrium is the same.

The distinction between $\widetilde{m}_t$, $\Delta\widehat{m}_t$, and $\widehat{m}_t$ is important because the former two are stationary when $\rho_z = 1$ while $\widehat{m}_t$ is not. Although one can arrive at the ST system by solving the DT model and re-defining variables even if $\rho_z = 1$, one should not use data analogue of $\widehat{m}_t$ in estimation because $\widehat{m}_t$ is a non-stationary model variable with $\rho_z = 1$. Recall that classical estimation assumes that the data are stationary and, thus, estimation requires stationary data analogues of the model concepts. On the other hand, $\widetilde{m}_t$ is stationary when $\rho_z = 1$ and thus is a model concept suitable for estimation.

2.3 Filtering the Data

To take the model to the data, one needs stationary data analogue of the model concepts. Suppose we observe the data for $d_t = (c_t, k_t, l_t)$. Let $d_t^c = (c_t^c, k_t^c, l_t^c) = (c_t - c_t^\tau,\, k_t - k_t^\tau,\, l_t - l_t^\tau)$ denote the data filtered to become stationary. We consider three possibilities.

- Linear Trend (LT):

$$c_t^c = c_t - \bar{g}t, \quad k_t^c = k_t - \bar{g}t, \quad l_t^c = l_t. \tag{4}$$

- HP Trend (HP):

$$c_t^c = HP(L)c_t, \quad k_t^c = HP(L)k_t, \quad l_t^c = l_t. \tag{5}$$

- First Difference (FD):[4]

$$c_t^c = \Delta c_t - \bar{g}, \quad k_t^c = \Delta k_t - \bar{g}, \quad l_t^c = \Delta l_t. \tag{6}$$

[3] One may also exploit cointegration relationships to construct stationary linear combinations of the non-stationary variables. For example, $c_t - y_t$ is stationary for all $|\rho_z| \leq 1$. That is, the cointegration vector (-1, 1) nullifies the deterministic and stochastic trends in c_t and y_t, if they exist. Although in our basic model using this vector in estimation does not bring in new information since $c_t - y_t \propto l_t$, cointegration vectors (e.g., $y_t - k_t$) can enrich the model and make it more robust to problems that can arise when ρ_z approaches one. This approach is exploited in Cogley (2001) when he estimates Euler equations for cointegrated variables and variables in growth rates.

[4] Even though the model predicts that labor is stationary, we first difference all series in the data because we solve the Δ^1DT model in first differences for all variables.

Ideally, linearly detrended data would replace the unobserved model variable $\widehat{m}_t$ when $|\rho_z| < 1$, while the HP filtered and first differenced data would stand in for $\widetilde{m}_t$ and $\Delta\widehat{m}_t$ when $\rho_z = 1$. HP filter can be and often is used in conjunction with $\widehat{m}_t$ when $|\rho_z| < 1$ because HP removes time trends as well. It is well known that the Hodrick-Prescott (HP) filter can alter the gain and phase of the cyclical components of the data (see e.g. Cogley and Nason (1995)) and change the error structure (see e.g. Singleton (1988)) in the univariate or single-equation framework. We examine formally how HP filter affects estimation in the DSGE context.

2.4 Estimation Procedure

Various non-Bayesian methods have been used to estimate the model as a system of equations. Two-step minimum distance approach (e.g., Sbordone (2006)), GMM/covariance structure (e.g., Christiano and den Haan (1996), Christiano and Eichenbaum (1992)), as well as simulation estimation (e.g., Altig et al. (2004)) can all be used. Ruge-Murcia (2005) provides a review of these methods. We use a methods of moments estimator that minimizes the distance between data moments and model-implied moments. Our estimation procedure can be summarized as follows:

1: Compute $\widehat{\Omega}^d(0) = cov(d_t^c)$, the covariance matrix of filtered series. Likewise, compute $\widehat{\Omega}^d(1)$, the first order sample auto-covariance.

2: Solve the rational expectations model (1), (2), or (3) for a guess of Θ, where Θ is the vector of structural parameters. Use the solution to compute $\Omega^m(0)$ and $\Omega^m(1)$, the model implied covariance and autocovariance matrix for model variables (which would be $\widehat{m}_t$, $\widetilde{m}_t$, or $\Delta\widehat{m}_t$).

3: Let $\widehat{\omega}^d = (vech(\Omega^d(0))' \;\; vech(\Omega^d(1))')'$ and let $\omega^m(\Theta) = (vech(\Omega^m(0))' \;\; vech(\Omega^m(1))')'$. Find structural parameters $\widehat{\Theta} = \text{argmin}_\Theta \left\|\widehat{\omega}^d - \omega^m(\Theta)\right\|$.[5]

Before estimation, a researcher needs to take a stand on two issues. First, he/she must decide whether the model is solved in terms of $\widehat{m}_t$, which is stationary under DT but not ST, or $\widetilde{m}_t$, which is stationary under ST. In doing so, the researcher is also making an assumption whether the shocks in the model are permanent or transitory. Second, the researcher needs to map the model variables (which are stationary) to the observed data and must decide how to filter the data. Problem (DD) arises when the two steps are not mutually consistent. Problem (MTS) arises when

[5] One can use the values of the ratios or other first moments contained in the constant terms as additional moments in estimation. However, since estimation of DSGE model is typically based on the second moments of the cyclical component of the data, we do not consider these additional moments in our analysis. As Cogley (2001) observes, "The RBC methodology is motivated by a desire to formulate estimators and tests that are not too distorted by trend misspecification." This separation of trend and cycle may be more problematic for general equilibrium models with endogenous growth and models without balanced growth path.

the data analogue of the model variables are not stationary or are over-differenced; that is, the assumed trend in the model is different from the trend in the data.

To illustrate the problems, consider the following combinations of model variables and data filtering techniques:

True Model	Assumed Model and Variables	Filtering	Problems
1. DT	DT, $\widehat{m}_t$	LT	-
2. DT	DT, $\widehat{m}_t$	HP	(DD)
3. ST	ST, $\Delta^1\widehat{m}_t$	FD	-
4. ST	ST, $\widetilde{m}_t$	HP	(DD)
5. DT	ST, $\widetilde{m}_t$	HP	(DD),(MTS)
6. ST	DT, $\widehat{m}_t$	LT	(MTS)

Of the six configurations, (1) and (3) are correctly specified and the data are appropriately filtered. In both cases, the assumed trend is identical to the trend in the data and, thus, there is no Problem (MTS). Because the researcher applies the same filter to the model variables and the data series, Problem (DD) is not an issue. In case (2), the assumed trend in the model is consistent with the actual trend in the data (both are deterministic time trends) and there is no Problem (MTS). However, the HP filter applied to the data series has different properties than the filter in the model, which is the linear time trend. Since these two filters do not agree in general, the researcher faces Problem (DD). A similar problem arises in case (4). In case (6), the assumed trend and the choice of detrending technique are consistent; that is, the researcher applies an appropriate filter given his or her assumption about the trend. Hence, Problem (DD) does not apply for this case. On the other hand, because the researcher has to choose either DT or ST before estimation, his or her choice of DT is not consistent with the true data generating process (ST) and, consequently, Problem (MTS) applies to this case. Likewise, in case (5), the choice of the trend in the model (DT) does not agree with the trend in the data (ST). In addition, the choice of the filtering technique in the data is not consistent with the assumed trend in the model. It follows that Problem (MTS) is further complicated by Problem (DD).

Two observations can be made. First, Problem (DD) involves only inconsistency between the model and the data trend. One can always circumvent the problem by applying the same filtering technique to the model variables and the data series. For example, the researcher can generate $\widehat{m}_t$, apply HP filter to generated $\widehat{m}_t$ and match the moments of filtered $\widehat{m}_t$ to the moments of HP filtered data. Although this procedure does not have Problem (DD), it is much slower and less efficient than alternative methods we describe below.

Second, Problem (MTS) arises only when the assumption about the trend in the model is different from the actual trend in the data, and this assumption has to be made before estimation.

A solution to Problem (MTS) is a flexible framework that nests DT and ST so that the researcher does not have to take a stand on whether $\rho_z < 1$ or $\rho_z = 1$. These two observations suggest that to address Problems (DD) and (MTS), the researcher needs an approach that *i*) transforms the data and model variables in the same way and *ii*) yields stationary series for all $|\rho_z| \leq 1$. In the next section, we develop such a framework.

2.5 A Robust Approach

In this subsection, we consider an approach that is robust to whether shocks are permanent or transitory. To begin, recall that the DT model solves the following system of equations:

$$E_t\Gamma_2^D \widehat{m}_{t+1} = \Gamma_0^D \widehat{m}_t + \Gamma_1^D \widehat{m}_{t-1} + \Psi_2^D u_{t+1}^z + \Psi_0^D u_t^z$$

where the Γ and Ψ matrices are defined in (1). We propose to solve an alternative representation of the same model. Let $\Delta^{\rho_z} = 1 - \rho_z L$ be the quasi-differencing operator. Then for a given ρ_z, a quasi-differenced representation of the DT model can be obtained by multiplying both sides of each equation in (1) by Δ^{ρ_z}:

The Quasi-Differenced DT Model

$$\begin{aligned} E_t \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & A_0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta^\rho \widehat{c}_{t+1} \\ \Delta^\rho \widehat{k}_{t+1} \\ \Delta^\rho \widehat{l}_{t+1} \end{bmatrix} &= \begin{bmatrix} -1 & 0 & -\alpha \\ 1 & A_0 & 0 \\ A_1 & A_2 & \alpha-1 \end{bmatrix} \begin{bmatrix} \Delta^\rho \widehat{c}_t \\ \Delta^\rho \widehat{k}_t \\ \Delta^\rho \widehat{l}_t \end{bmatrix} + \begin{bmatrix} 0 & \alpha & 0 \\ 0 & 0 & 0 \\ 0 & A_4 & 0 \end{bmatrix} \begin{bmatrix} \Delta^\rho \widehat{c}_{t-1} \\ \Delta^\rho \widehat{k}_{t-1} \\ \Delta^\rho \widehat{l}_{t-1} \end{bmatrix} \\ &+ \begin{bmatrix} 0 \\ -A_0 \\ 0 \end{bmatrix} \Delta^\rho u_{t+1}^z + \begin{bmatrix} 1-\alpha \\ 0 \\ \alpha-1 \end{bmatrix} \Delta^\rho u_t^z. \end{aligned} \tag{7}$$

Since $u_t^z = \rho^z u_{t-1}^z + e_t^z$, we have

$$E_t\Gamma_2^D \Delta^{\rho_z} \widehat{m}_{t+1} = \Gamma_0^D \Delta^{\rho_z} \widehat{m}_t + \Gamma_1^D \Delta^{\rho_z} \widehat{m}_{t-1} + \Psi_1^D e_{t+1}^z + \Psi_0^D e_t^z$$

where $\Delta^{\rho_z} \widehat{m}_t = (\Delta^{\rho_z} c_t, \Delta^{\rho_z} k_t, \Delta^{\rho_z} l_t)$. Note that the error term in (7) is an i.i.d. innovation and therefore $\Delta^{\rho_z} \widehat{m}_t$ is stationary for all $|\rho_z| \leq 1$. The appeal of the quasi-differenced representation is that it is valid for all ρ_z less than or equal to one; (7) is just a special case of (2) at $\rho_z = 1$. Partition $\Theta = (\Theta^-, \rho_z)$. The deep parameters can be estimated as follows:

<u>The Quasi-Differenced (QD) Estimator:</u> Initialize ρ_z.

1: Quasi difference the observed data at ρ_z to obtain

$$c_t^c = \Delta^{\rho_z}(c_t - \bar{g}t), \quad k_t^c = \Delta^{\rho_z}(k_t - \bar{g}t), \quad l_t^c = \Delta^{\rho_z} l_t, \tag{8}$$

and let $\Delta_z^{\rho} d_t^c = (c_t^c, k_t^c, l_t^c)$.[6]

2: Compute $\widehat{\Omega}^d_{\Delta^{\rho_z}}(0) = cov(\Delta^{\rho_z} d_t^c)$, the covariance matrix of $\Delta^{\rho_z} d_t^c$, and the autocovariance matrix $\widehat{\Omega}^d_{\Delta \rho_z}(1)$. Define $\widehat{\omega}^d_{\Delta^{\rho_z}} = (vech(\Omega^d_{\Delta^{\rho_z}}(0))' \;\; vech(\Omega^d_{\Delta^{\rho_z}}(1))')'$;

3: For a given ρ_z and Θ^-, solve (7) and compute $\Omega^m_{\Delta^{\rho_z}}(0)$ and $\widetilde{\Omega}^m_{\Delta^{\rho_z}}(1)$, the model implied covariance and autocovariance matrices. Define $\widehat{\omega}^m_{\Delta^{\rho_z}} = (vech(\Omega^m_{\Delta^{\rho_z}}(0))' \;\; vech(\Omega^m_{\Delta^{\rho_z}}(1))')'$;

4: Find the structural parameters $\widehat{\Theta} = \arg\min_{\Theta} \left\| \widehat{\omega}^d_{\Delta^{\rho_z}} - \omega^m_{\Delta^{\rho_z}}(\Theta) \right\|$.

Note that ρ_z and Θ^- are estimated simultaneously. The quasi-differenced estimator differs from the covariance estimator of the previous section in one important respect. The parameter ρ_z now affects both the moments of the model and the data since the latter are computed for the quasi-transformed data. Conceptually, the crucial feature is that the quasi-transformed data are stationary irrespective of ρ_z. Thus, the QD estimator resolves Problem (DD) by applying the same transformation (filter) to the data and model and tackles Problem (MTS) by using a transformation that yields stationary series for any $|\rho_z| \leq 1$. Because $\Delta^{\rho_z} m_t$ is stationary, the estimation problem can be studied under the assumptions of extremum estimation. Under regularity conditions, standard $\sqrt{T}$ asymptotic normality results hold (see Section 3.4).

At this point it is useful to highlight the difference between our and Cogley's (2001) approaches. Although both methods do not require the researcher to take a stand on the properties of the trend dynamics before estimation, our approach has several advantages. First, quasi-differencing can easily handle multiple I(1) or highly persistent shocks. In contrast, using cointegration relationships works only for certain types of shocks. For example, if the shock to disutility of labor supply is an I(1) process, there is no cointegration vector to nullify a trend in hours. Second, cointegration often involves estimating identities and therefore the researcher has to add an error term (typically measurement error) to avoid singularity. Our approach does not estimate specific equations and hence does not need to augment the model with additional, atheoretical shocks. Finally, using unconditional cointegration vectors may make estimation of some structural parameters impossible. For instance, the parameters governing short-run dynamics such as adjustment costs may be not estimated in this setting because the term arising due to adjustment costs is zero on average by construction (i.e., adjustment cost is typically zero in steady state). In contrast, our approach utilizes short-run dynamics in estimation and thus can estimate the parameters affecting short-run dynamics of the variables. Overall, our approach can be used in a broader array of situations and we exploit different properties of the model in estimation.

[6] Since projecting series on linear trend yields super-consistent estimates of the coefficient on the time trend, one can ignore the error induced by removing the linear time trend when he or she applies standard asymptotic inference.

Naturally, if elements of $\Delta^{\rho_z}\widehat{m}_t$ are stationary concepts when $|\rho_z| \leq 1$, they are also stationary when the data are quasi differenced at $\rho_z = 1$. This suggests that solving the first difference representation of the DT model will also yield robust estimates:

ΔDT Model

$$E_t \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & A_0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta\widehat{c}_{t+1} \\ \Delta\widehat{k}_{t+1} \\ \Delta\widehat{l}_{t+1} \end{bmatrix} = \begin{bmatrix} -1 & 0 & -\alpha \\ 1 & A_0 & 0 \\ A_1 & A_2 & \alpha-1 \end{bmatrix} \begin{bmatrix} \Delta\widehat{c}_t \\ \Delta\widehat{k}_t \\ \Delta\widehat{l}_t \end{bmatrix} + \begin{bmatrix} 0 & \alpha & 0 \\ 0 & 0 & 0 \\ 0 & A_4 & 0 \end{bmatrix} \begin{bmatrix} \Delta\widehat{c}_{t-1} \\ \Delta\widehat{k}_{t-1} \\ \Delta\widehat{l}_{t-1} \end{bmatrix}$$
$$+ \begin{bmatrix} 0 \\ -A_0 \\ 0 \end{bmatrix} \Delta u^z_{t+1} + \begin{bmatrix} 1-\alpha \\ 0 \\ \alpha-1 \end{bmatrix} \Delta u^z_t. \tag{9}$$

Observe that when $\Delta u^z_t = (\rho_z - 1)u^z_{t-1} + e^z_t$, and $\rho_z < 1$, ρ^z remains a parameter of the model (3) unless it is constrained to be one. To stress that ρ_z is a free parameter and contrast it with the constrained specification, we do not put a superscript on the first difference operator. The difference between the constrained $\Delta^1 DT$ and unconstrained ΔDT models is that the unconstrained model is valid whether or not $\rho_z = 1$, while the constrained model is an alternative representation of the ST model and is thus valid when $\rho_z = 1$. Note that the QD estimator and ΔDT estimator are equivalent when $\rho_z = 1$. While the moments of ΔDT model are robust to whether ρ_z is on the unit circle, it is less efficient relative to the quasi-differenced model since the data will be over-differenced when the data are already stationary. The estimation procedure for the unconstrained ΔDT estimator is as follows:

1: First difference the observed data to obtain

$$c^c_t = \Delta c_t - \bar{g}, \quad k^c_t = \Delta k_t - \bar{g}, \quad l^c_t = \Delta l_t. \tag{10}$$

and let $\Delta d^c_t = (c^c_t, k^c_t, l^c_t)$.

2: Compute $\widehat{\Omega}^d_\Delta(0) = cov(\Delta d^c_t)$, the covariance matrix of Δd^c_t, and the autocovariance matrix $\widehat{\Omega}^d_\Delta(1)$. Define $\widehat{\omega}^d_\Delta = (vech(\Omega^d_\Delta(0))' \;\; vech(\Omega^d_\Delta(1))')'$;

3: For a given Θ, solve (9) and compute $\Omega^m_\Delta(0)$ and $\widetilde{\Omega}^m_\Delta(1)$, the model implied covariance and autocovariance matrices. Define $\widehat{\omega}^m_\Delta = (vech(\Omega^m_\Delta(0))' \;\; vech(\Omega^m_\Delta(1))')'$;

4: Find the structural parameters $\widehat{\Theta} = \arg\min_\Theta \left\| \widehat{\omega}^d_\Delta - \omega^m_\Delta(\Theta) \right\|$.

The two robust methods can be summarized as follows:

True Model	Assumed Model/ Variables	Data	Estimator
ST, DT	QD, $\Delta^{\rho_z}\widehat{m}_t$	$\Delta^{\rho_z}d_t^c$	QD
ST, DT	ΔDT, $\Delta\widehat{m}_t$	Δd_t^c	ΔDT.

Note that these methods do not require the researcher to take a stand on whether $\rho_z < 1$ or $\rho_z = 1$ before estimation because these two cases are nested within QD and ΔDT framework.

3 Simulations

3.1 Setup and Calibration

We generate the data as either DT (deterministic trends) or ST (stochastic trends) using the model equations for the stationary (i.e., normalized) variables. The model variables are then rescaled back to non-stationary form and treated as observed data $d_t = (c_t, k_t, y_t, l_t)$ that the researcher takes as given. The researcher then decides (i) whether to use the model equations implied by DT or ST for estimation, and (ii) how to detrend the data.

We estimate $\Theta = (\alpha, \rho, \sigma)$ and treat parameters $(\beta, \delta, \theta, \bar{g})$ as known. We calibrate the model as follows: capital intensity $\alpha = 0.33$; disutility of labor $\theta = 1$; discount factor $\beta = 0.99$; depreciation rate $\delta = 0.1$; gross growth rate in technology $\bar{g} = \bar{\gamma} = 1.005$. We restrict the admissible range of the estimates of α to [0.01,0.99]. We vary the persistence of shocks to technology u_t^z. The parameter ρ_z takes values (0.5, 0.95, 0.99, 1). The admissible range for $\widehat{\rho}_z$ in the DT model is [-0.999,0.999]. Since for now we have only one shock in the model, we set the standard deviation of e_t^z to $\sigma = 0.1$ without loss of generality. We perform 500 replications for each choice of parameter values. For each replication, we create series with T=300 observations which is a typical sample size in applied macroeconometric analysis.

In the simulations, we use a covariance structure estimator that minimizes the distance between the observed unconditional autocovariances of the data and those implied by the model. We described the estimator in the previous sections. To minimize distortions associated with poor estimation of the optimal weighting matrix, we use an identity weighting matrix in our method of moments estimator. In all simulations and for all estimators, we set starting values in optimization routines equal to the true parameter values.

3.2 Results for the Baseline Model

We report simulation results for the baseline one-shock model in Table 2 and present the kernel density estimates for parameter estimates in Figure 1. We use the following notation to label difference cases. In case (XX,YY), XX stands for the method used to filter the data, while YY stands for the assumed model. Thus, (LT,$\widehat{m}_t$) means that the data used in estimation are residuals from projection on a time trend, and the assumed model is expressed in terms of $\widehat{m}_t$ with $|\rho_z| < 1$. The DGP is given in the first column.

Our simulations suggest that combinations (QD,$\Delta^{\rho_z}\widehat{m}_t$) and (FD,$\Delta\widehat{m}_t$), which are reported in columns 4 and 6 and correspond to the QD and unconstrained ΔDT estimators respectively, yield estimates generally centered at the true values. The distribution of the estimates is bell-shaped and well-behaved uniformly for all values of ρ_z. That is, the performance of QD and ΔDT estimators is not affected as ρ_z approaches one. This pattern is recurrent in all simulations. In contrast, other estimators exhibit significant biases and larger dispersion of estimates especially when ρ_z is close to a unit circle. Below we document their properties and explain why these estimators tend to underperform.

Consider first the (LT,$\widehat{m}_t$) combination when the researcher uses series after linear detrending as the data concept and $\widehat{m}_t$ as the model concept of the observed variables (column 1, Table 2). For small to moderate values of ρ_z, this combination performs well: the distribution of the parameter estimates is centered at true values. However, as ρ_z increases the performance of the (LT,$\widehat{m}_t$) combination quickly deteriorates. There is a significant upward bias in the estimates of the capital share α. Furthermore, this bias increases with ρ_z so much that at $\rho_z = 1$, the mean of $\widehat{\alpha}$ is close to one. The bias in $\widehat{\sigma}$ also worsens rapidly as ρ_z approaches one and the dispersion of the estimates is large as seen from the flat density of $\widehat{\sigma}$ in Figure 1. The estimates of ρ_z tend to be relatively close to true values up to $\rho_z = 0.95$. As ρ_z approaches one, however, there is a strong downward bias in $\widehat{\rho}_z$. For example at $\rho_z = 1$, the mean of $\widehat{\rho}_z$ is approximately 0.7.

Note that the (LT,$\widehat{m}_t$) case can not only significantly bias the estimates but it can also yield multi-modal distribution of the estimates. For example, the case with $\rho_z = 0.99$ has two peaks in the distribution of $\widehat{\alpha}, \widehat{\sigma}$ and $\widehat{\rho}_z$, i.e., the objective function of the covariance structure estimator has two or more local optima. This observation is particularly troubling for users of standard optimization routines as these routines can fail to escape from local optima. Importantly, estimates based on different local optima can lead to drastic changes in the economic interpretation of the estimates.

The case of $\rho_z = 1$ (last row of the (LT,$\widehat{m}_t$) column) is particularly interesting because linear detrending is commonly used in estimation of DSGE models, as seen from Table 1. Projecting series with a unit root on time trend is known to lead to spurious cycles in univariate analysis (e.g. Nelson and Kang (1981)). Our results suggest that in systems estimation such as the one considered here, linear detrending leads to extremely strong biases in the estimates of the structural parameters. Since technology shocks appear to be highly persistent and well approximated with unit root (Problem MTS), researchers should be very cautious with using linearly detrended data for estimation of DSGE models in applied work.

Turning to the (HP,$\widehat{m}_t$) combination in column 2, the estimates of ρ_z have a strong downward bias. On the other hand, there is a strong upward bias in $\widehat{\alpha}$ and $\widehat{\sigma}$.[7] These estimates suggest larger but less persistent shocks to technology as well as a significant role of capital as a mechanism for propagating shocks in the model. To understand this pattern, recall that HP filter removes not only the linear trend but also low frequency variation in the series. When ρ_z is large, HP filter can significantly alter the properties of the series. More specifically, HP filter changes not only the persistence of the series (recall Cogley and Nason (1995)) but also the relative volatility and serial correlation of the series(see e.g. King and Rebelo (1993) and Harvey and Jaeger (n.d.)). This translates into biased estimates of all parameters because the estimator is forced to match the properties of the altered data which is different from the model concept of observed variables.

Under (HP,$\widetilde{m}_t$), ρ_z is fixed at 1 and the model variables are $\widetilde{m}_t$. As seen from column 3, the estimates of α and σ remain unsatisfactory. The estimates of α's are lumped at the boundary of the admissible range [0.01, 0.99] and the estimates of σ tend to be very close to zero. In other words, the estimated model suggests that shocks to technology are very small but propagation through capital accumulation is strong. Why does this happen? Note that in the ST model defined in terms of $\widetilde{m}_t$, the dynamics of the variables tend to have weak serial correlation because deviations from u_t are transitory and quickly dissipating as variables such as consumption adjust to almost full strength in response to permanent shocks to technology. On the other hand, HP filter leaves out sizable serial correlation in the filtered data. Thus, the fitted model is forced to produce parameter values that have strong propagation mechanism to generate relatively strong serial correlation in deviations from the stochastic trend.

To understand the strong downward bias in $\widehat{\sigma}$, note that $var(\widetilde{y}_t) < var(\widetilde{c}_t)$ and $var(\widetilde{k}_t) > var(\widetilde{c}_t)$ in the model, while $var(y_t^c) < var(c_t^c)$ and $var(k_t^c) < var(c_t^c)$ in the HP-filter series. As α approaches one, $var(\widetilde{y}_t)$ and $var(\widetilde{k}_t)$ become approximately equal to $var(\widetilde{c}_t)$ and thus the gap in the relative volatility between output, capital and consumption resembles the relative volatility in the data. At the same time, larger values of α increase the volatility of the series and the estimator decreases the size of the shocks to match the level of volatility in the data. Hence, there is a strong downward bias in $\widehat{\sigma}$. Overall, results for combinations (HP,$\widehat{m}_t$) and (HP,$\widetilde{m}_t$) suggest that Problems (DD) and (MTS) can significantly affect estimates of structural parameters and can lead to erroneous economic interpretations.

To get a sense of how much difference filtering can make, consider the combination (FD,$\Delta^1 m_t$),

[7] Note that we do not HP-filter labor series as labor is stationary irrespective of whether $\rho_z < 1$ or $\rho_z = 1$. Results do not change qualitatively when we estimate the model using HP-filter labor series.

reported in column 5, Table 2. It performs reasonably well when $\rho_z \approx 1$, that is, ST is the correct assumption and first differencing is correctly applied to data and model variables. As ρ_z departs from one, Problem (MTS) manifests in an increasing upward bias in both $\widehat{\alpha}$ and especially $\widehat{\sigma}$. Note that despite the fact that estimates based on (FD,$\Delta^1 m_t$) exhibit sizable biases when ρ_z moves away from one, (FD,$\Delta^1 m_t$) dominates (HP,$\widetilde{m}_t$) by a large margin. This pattern is typical in our simulations.

3.3 Estimation of Propagation Mechanisms

Clearly, the absurdly large estimates of α or similar problems with the estimates of deep parameters can alert the researcher that the model is likely misspecified and he or she must make adjustments to the model. One possible and popular adjustment is to introduce serial correlation in the growth rates of structural shocks such as technology. Interestingly, when we introduce such correlation in the growth rates of technology (not reported) and estimate the model using (HP,$\widetilde{m}_t$) combination, the estimates of α take more plausible values in the range of 0.4-0.5. However, this modification in the model is ad hoc and more importantly it indicates that improper choice of filtering techniques and model concepts can induce the researcher to augment correctly specified models with spurious mechanisms of propagation and amplification to match the moments of the data.[8]

To highlight this point, we augment the basic model with internal habit in consumption. This modification is a popular way to introduce greater persistence and amplification in business cycle models. Specifically, consider an alternative utility function:

$$\max \sum \beta^t \left[\ln(C_t - \phi C_{t-1}) - L_t \right]$$

where $\phi \in [-0.999, 0.999]$ measures the degree of habit in consumption. In this model, the researcher estimates $(\alpha, \phi, \rho, \sigma)$. We set $\phi = 0.8$ to investigate how the treatment of the trends affect estimates of internal propagation mechanisms as well as estimates of other structural parameters. We report results in Table 3 and Figure 2. To save space, we do not consider the case of $\rho_z = 0.5$ and present kernel densities of the estimates for only the case of $\rho_z = 0.95$ in Figure 2.

Similar to the results in the previous section, QD and ΔDT perform well. The bias in the estimates is generally negligible and the distribution of the estimates is well-behaved. Overall, QD and ΔDT strongly dominate alternative estimators whose performance we examine below.

The combination (LT,$\widehat{m}_t$) has a relatively small upward bias in $\widehat{\phi}$ when $\rho_z = 0.95$ but the performance of (LT,$\widehat{m}_t$) quickly deteriorates as ρ_z approaches one, see column 1 of Table 3. Specifically, at $\rho_z = 0.99$ the mean value of $\widehat{\phi}$ is close to the the true value of ϕ but the dispersion of $\widehat{\phi}$

[8]For example, Doorn (2006) shows in simulations that HP filtering can significantly alter the parameter estimates governing dynamic properties in his inventory model.

rapidly increases indicating that the distribution of $\widehat{\phi}$ is quite flat. When $\rho_z = 1$, the mean of $\widehat{\phi}$ sharply drops and the dispersion of $\widehat{\phi}$ increases further. In fact, the kernel density of $\widehat{\phi}$ is practically flat (not reported) so that researcher using (LT,$\widehat{m}_t$) may end up with effectively any estimate of ϕ. Note that introducing habit formation changes the pattern of biases in the estimates of other parameters when compared to the baseline model without habit formation. In particular, although $\widehat{\alpha}$ is upwardly biased in the model with and without habit formation, there is a downward bias in $\widehat{\rho}_z$ and $\widehat{\sigma}$ for the model with habit formation which is different from the results for the baseline model without habit formation. Note that it is very hard to predict the sign of the bias in general. Small modifications in a model can lead to distortionary effects reinforcing or attenuating each other so that estimates can over- or understate the magnitudes of structural parameters. The direction of the bias is highly model specific and a priori ambiguous.

Under the (HP,$\widehat{m}_t$) combination, when the researcher uses HP filter to remove the trend, there are larger distortions to $\widehat{\phi}$. The estimate of ϕ has a clear upward bias when $\rho_z = 0.95$. However, the mean value of $\widehat{\phi}$ understates the degree of the bias as the distribution of $\widehat{\phi}$ has a thick left tail. As ρ_z approaches one, the dispersion of $\widehat{\phi}$ increases dramatically, which is a manifestation of the flat distribution of $\widehat{\phi}$. This finding suggests that identification of ϕ from the filtered data may be poor. Indeed, identification of ϕ comes from low frequency variation in the data but this frequency is removed or greatly attenuated by the HP filter. Other estimates are also biased. In particular, $\widehat{\alpha}$ has a stronger upward bias than in the case without habit formation. The bias in $\widehat{\sigma}$ decreases with ρ_z while it increases with ρ_z in the model without habit formation. Also note that the extent of the bias in $\widehat{\rho}_z$ is smaller in this model than in the model without habit formation.

The (HP,$\widetilde{m}_t$) combination which imposes $\rho_z = 1$ tends to produce results similar to the previous case but with more acute identification problems for ϕ as the density of $\widehat{\phi}$ is fairly flat even for $\rho_z = 0.95$. Also note that in this case, there is a somewhat smaller pressure on $\widehat{\alpha}$ and the mean of $\widehat{\alpha}$ is away from the boundary of the admissible space for the estimate of α because part of the adjustment to match the data moments is absorbed by changes in the estimates of ϕ. Interestingly, as ϕ increases towards one, the bias in $\widehat{\alpha}$ turns from upward to downward. Again, note that using an alternative combination (FD,$\Delta^1 m_t$) can greatly improve the estimates when $\rho_z \approx 1$, which is similar to the baseline case without habit formation.

3.4 Statistical Properties

The above simulations indicate that our quasi-differenced estimates are close to the true value and the difference between the estimates and the true value is by and large symmetrically distributed. Note that our quasi-differenced estimator is nothing but a non-linear GMM estimator using an identity as weighting matrix and stationary (after quasi-differencing) data. Since our estimator satisfies prerequisites for standard asymptotic inference, one may expect that the t-statistic for estimates based (QD,$\Delta^{\rho_z}\widehat{m}_t$) (or (FD,$\Delta\widehat{m}_t$)) may be well approximated by the normal distribution for large T and our estimator is $\sqrt{T}$ consistent and asymptotically normal. In this subsection, we assess this conjecture.

Let $v_t^c = (d_t^c, d_{t-1}^c, \ldots, d_{t-s}^c)$ be the vector of stacked data so that $\omega^d = vech(cov(v_t^c))$. Likewise, define v_t^m, the vector of stacked model variables, and $\omega^m = vech(cov(v_t^m))$. It is convenient to define

$$\frac{1}{T}\sum_{t=1}^{T} g_t = \bar{g} = \omega^d - \omega^m.$$

By construction of the quasi-differenced data, $\bar{g}$ is the sample mean of a stationary ergodic process and $\sqrt{T}\bar{g}\xrightarrow{d}N(0,S)$. By assumption, g_t is continuous in Θ, which is a compact. The assumptions of the GMM estimator such as stated in Hayashi (2000) apply. Let G be the matrix of derivatives of g with respect to Θ. Then

$$\widehat{\Theta} = \underset{\Theta}{\operatorname{argmin}}\, J = T \times \bar{g}'\bar{g}$$

yields

$$\sqrt{T}(\widehat{\Theta} - \Theta)\xrightarrow{d}N(0, (G'G)^{-1}G'S'G(G'G)^{-1}).$$

We employ Newey-West estimator of S, i.e., $S = T^{-1}(v_t^{c'}v_t^c) + \sum(1 - j/(p+1))T^{-1}(v_t^{c'}v_{t-j}^c)$. Using this standard machinery of asymptotic inference, we compute t-statistic for the parameters of the baseline one-shock model. If the inference is $\sqrt{T}$ consistent, the simulated density for t-statistic should be close to the p.d.f. of the standard normal random variable. We focus on three combinations (QD,$\Delta^{\rho_z}\widehat{m}_t$), (FD,$\Delta\widehat{m}_t$), and (LT,$\widehat{m}_t$) and report the kernel density of t-statistic for T=300 and T=2,000 in Figures 3 and 4.

The figures show that the distribution based on the QD estimator for α and σ is generally close to the N(0,1) density. Likewise, apart from the case when $\rho_z = 1$, the distribution of the t-statistic for $\widehat{\rho}_z$ is also closely approximated by the standard normal distribution.[9] The distribution of t-statistic based on ΔDT is somewhat less impressive but nonetheless it is a much better approximation to N(0,1) than the approximation provided by a more commonly used (LT,$\widehat{m}_t$) combination.

While $\sqrt{T}$ asymptotic normality of α and σ may not seem surprising, $\sqrt{T}$ consistency and asymptotic normality of ρ_z as suggested by the simulations may be unexpected. This is because

[9] Note that the model does not have a unique rational expectations equilibrium if $\rho_z > 1$. Since we consider only parameter values consistent with the existence of a unique rational expectations equilibrium, the estimate for ρ_z becomes distorted as $\widehat{\rho}_z$ cannot exceed unity. However, in the linear single-equation case considered in Gorodnichenko and Ng (2007), the distribution of $\widehat{\rho}_z$ is well behaved and close to the standard normal since $\widehat{\rho}_z$ is allowed to exceed one.

almost all the results in the literature on integrated data suggest that estimates of the largest autoregressive root tend to be super consistent with a convergence rate of T, but that the asymptotic distribution is highly skewed. This is at odds with our finding that the distribution of $\widehat{\rho}_z$ appears symmetric. This surprising and curious result was investigated in more detail in Gorodnichenko and Ng (2007) in the much simpler case of estimating the largest autoregressive root of an autoregressive process using a covariance structure estimator such as the one considered in this paper. In a nutshell, there are two aspects of the estimator that leads to this result. First, the covariance estimator is non-linear in the parameter of interest, and second the covariance is defined for the error process rather than the data itself. It was shown in our companion paper that this particular non-linear estimator has an asymptotic distribution that is continuous at the unit circle is conditionally normal, which greatly facilitates inference. The cost, which seems warranted, is that the estimator looses its super-consistent property, though it continues to be $\sqrt{T}$ consistent. The estimator considered in this paper is different in that it is a systems estimator, but it appears to have the same desirable properties as the estimator considered in Gorodnichenko and Ng (2007).

4 The General Formulation

Our Monte Carlo experiments suggest that the (QD,$\Delta^{\rho_z}\widehat{m}_t$) and (FD,$\Delta\widehat{m}_t$) estimators outperform popular alternatives in terms of providing less dispersed and biased estimates of structural parameters. Combinations (QD,$\Delta^{\rho_z}\widehat{m}_t$) and (FD,$\Delta\widehat{m}_t$) are similar in terms of having little or no bias. However, (QD,$\Delta^{\rho_z}\widehat{m}_t$) tends to have smaller dispersion and more well behaved distribution of the estimates than (FD,$\Delta\widehat{m}_t$) when at least one of the shocks is persistent.

The key to robustness is that these estimates do not depend on whether the data are persistent and that they apply basic transformations to both possibly persistent data and model variables so that there is a coherent mapping between the model and data. Effectively, Problems (DD) and (MTS) are addressed simultaneously because *i*) the researcher does not have to take a stand on the properties of the forcing variables (e.g., technology has a permanent or transitory shocks) as the transformed data and model variables are stationary for all parameter values describing persistence of forcing variables; and *ii*) the researcher applies the same transformation (filter) to model and data to make variables stationary so that filtered variables in the model and data have the same connotation. In other words, (QD,$\Delta^{\rho_z}\widehat{m}_t$) and (FD,$\Delta\widehat{m}_t$) preserve consistency between the data and model concepts irrespective of whether variables are stationary or not.

Our framework straightforwardly extends to more general cases. Suppose there are J shock processes u_{jt}, $j = 1, \ldots J$, and

$$(1 - \rho_j L)u_{jt} = e_{jt}, \quad j = 1, \ldots J$$

and some of the ρ_j may be on the unit circle. Let m_t be a vector of (predetermined, non-predetermined, plus exogenous) variables of the model and let $\widehat{m}_t$ be the vector of zero-mean variables that are deviations of m_t from the steady state values. While $\widehat{m}_t$ is demeaned, it is not normalized by technology or other non-stationary forcing variables. Normalization of $\widehat{m}_t$ results in $\widetilde{m}_t = \widehat{m}_t - u_t$ and is necessary to make model variables stationary when some shocks are permanent. As seen from Table 1, most studies assume that shocks are transitory and solve

$$\Gamma_0^D \widehat{m}_t = \Gamma_1^D E_t \widehat{m}_{t+1} + \Gamma_2^D \widehat{m}_{t-1} + \Psi_1^D u_{t+1} + \Psi_0^D u_t$$

A smaller number of studies assume stochastic trends and solve

$$\Gamma_0^S \Delta\widehat{m}_t = \Gamma_1^S E_t \Delta\widehat{m}_{t+1} + \Gamma_2^S \Delta\widehat{m}_{t-1} + \Psi_1^S e_{t+1} + \Psi_0^S e_t.$$

As illustrated in section 3, the estimation approach depends on what is the assumed model, how the observed data are filtered to become stationary. Let d_t be a vector of r observed variables. Partition the model variables $\widehat{m}_t = (\widehat{m}_{0t}, \widehat{m}_{1t})$ to have q_0 stationary and q_1 non-stationary state variables, denoted $\widehat{m}_{0t}$ and $\widehat{m}_{1t}$, respectively. The general solution in state space representation is

$$d_t = \delta_0 + \delta_1 \, t + B_0 \widehat{m}_{0t} + B_1 \widehat{m}_{1t} \tag{11a}$$

$$\widehat{m}_{0t} = \pi_0 \widehat{m}_{0t-1} + \Pi_0 u_t \tag{11b}$$

$$\widehat{m}_{1t} = \widehat{m}_{1t-1} + \Pi_1 e_t. \tag{11c}$$

The measurement equation (11a) links the $r \times 1$ observed variables m_t to the q model variables, $\widehat{m}_t$ through the matrices B_0 $(r \times q_0)$ and B_1 $(r \times q_1)$.[10] The parameters δ_0 and δ_1 are $r \times 1$ vectors of restricted constants to be estimated along with the other parameters. This ensures that the data are detrended using model consistent parameters. An alternative, used in Ireland (2004a) and many others, is to linearly detrend the data prior to estimation. This amounts to not imposing constraints on δ_0 and δ_1 to take on values implied by the model. If the model is correctly specified for the data, both methods of detrending are asymptotically equivalent. Without loss of generality, simply let $d_t^c = d_t - \delta_0 - \delta_1 t$ be the detrended data. When $q_1 = 0$ and $\widehat{m}_{0t} = \widehat{m}_t$, the measurement equation becomes

$$d_t^c = d_t - \delta_0 - \delta_1 t = B_0 \, \widehat{m}_t. \tag{12}$$

Equation (11b) is the law of motion for the state variables of a trend stationary model. The state variables with a unit root evolve according to (11c). When $q_0 = 0$, (11b) is irrelevant and the

[10] A vector of r measurement errors η_t can be added to the measurement equation as in Edge et al. (2005).

measurement equation is either

$$d_t^c = \Delta d_t - \delta_1 = B_1 \Delta \widehat{m}_t, \tag{13}$$

or

$$d_t^c = HP(L) d_t = B_1 \Delta \widetilde{m}_t. \tag{14}$$

Our quasi-differencing approach can be extended to the case of multiple shocks by defining

$$\Delta^\rho(L) = \prod_{j=1}^{J} (1 - \rho_j L).$$

In other words, the quasi-differencing operator is now the product of the J polynomials in lag operator that describes the dynamics of the J shocks. Then the J-shock quasi-differenced model is defined as

$$E_t \Gamma_2^D \Delta^\rho \widehat{m}_{t+1} = \Gamma_0^D \Delta^\rho \widehat{m}_t + \Gamma_1^D \Delta^\rho \widehat{m}_{t-1} + \Psi_1^D \Delta^\rho u_{t+1} + \Psi_0^D \Delta^\rho u_t, \tag{15}$$

The link between the data and the model is given by

$$\Delta^\rho (d_t - \delta_0 - \delta_1 t) \quad = \quad \Delta^\rho \widehat{m}_t.$$

This link is valid whether the true model is DT or ST. Note that one does not have to work with the product of $(1 - \rho L)$ operators for each shock. Following the insight from the comparison of the QD and unconstrained ΔDT estimators, one can use quasi-differencing only for shocks that are expected to be persistent and do not need to transform the data further to accommodate other, known-to-be stationary shocks. For example, if one knows that shocks to tastes dissipate quickly while technology shocks z_t are close to unit root, the researcher can use only $(1 - \rho_z L)$ in the Δ^ρ operator.

4.1 A Two Shock Example

To illustrate the multiple shock case, we re-introduce shocks to hours in the model so that the system is given by

$$\begin{aligned} E_t \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & A_0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \widehat{c}_{t+1} \\ \widehat{k}_{t+1} \\ \widehat{l}_{t+1} \end{bmatrix} &= \begin{bmatrix} -1 & 0 & -\alpha \\ 1 & A_0 & 0 \\ A_1 & A_2 & \alpha - 1 \end{bmatrix} \begin{bmatrix} \widehat{c}_t \\ \widehat{k}_t \\ \widehat{l}_t \end{bmatrix} + \begin{bmatrix} 0 & \alpha & 0 \\ 0 & 0 & 0 \\ 0 & A_4 & 0 \end{bmatrix} \begin{bmatrix} \widehat{c}_{t-1} \\ \widehat{k}_{t-1} \\ \widehat{l}_{t-1} \end{bmatrix} \\ &+ \begin{bmatrix} 0 & 0 \\ -A_0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_{t+1}^z \\ u_{t+1}^q \end{bmatrix} + \begin{bmatrix} 1 - \alpha & 1 \\ 0 & 0 \\ \alpha - 1 & 0 \end{bmatrix} \begin{bmatrix} u_t^z \\ u_t^q \end{bmatrix} \end{aligned} \tag{16}$$

Following the procedures we describe above, it is straightforward to write this model in terms of stationary variables $\widetilde{m}_t$ or Δm_t when shocks to technology or hours contain a unit root.

The QD approach extends naturally to the case of two shocks. Let $\Delta^\rho = (1-\rho^z L)(1-\rho^q L)$. The quasi-differenced representation of the two-shock model (16) is:

$$
\begin{aligned}
E_t \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & A_0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta^\rho \widehat{c}_{t+1} \\ \Delta^\rho \widehat{k}_{t+1} \\ \Delta^\rho \widehat{l}_{t+1} \end{bmatrix} \quad = \quad & \begin{bmatrix} -1 & 0 & -\alpha \\ 1 & A_0 & 0 \\ A_1 & A_2 & \alpha - 1 \end{bmatrix} \begin{bmatrix} \Delta^\rho \widehat{c}_t \\ \Delta^\rho \widehat{k}_t \\ \Delta^\rho \widehat{l}_t \end{bmatrix} + \begin{bmatrix} 0 & \alpha & 0 \\ 0 & 0 & 0 \\ 0 & A_4 & 0 \end{bmatrix} \begin{bmatrix} \Delta^\rho \widehat{c}_{t-1} \\ \Delta^\rho \widehat{k}_{t-1} \\ \Delta^\rho \widehat{l}_{t-1} \end{bmatrix} \\
& + \begin{bmatrix} 0 & 0 \\ -A_0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} e^z_{t+1} \\ e^q_{t+1} \end{bmatrix} + \begin{bmatrix} 1-\alpha & 1 \\ \rho^q A_0 & 0 \\ \alpha - 1 & 0 \end{bmatrix} \begin{bmatrix} e^z_t \\ e^z_t \end{bmatrix} \\
& + \begin{bmatrix} -\rho^q(1-\alpha) & -\rho^z \\ 0 & 0 \\ \rho^q(\alpha-1) & 0 \end{bmatrix} \begin{bmatrix} e^z_{t-1} \\ e^q_{t-1} \end{bmatrix}.
\end{aligned}
$$

This representation is valid when none, one or both shocks are non-stationary. It is easy to see that when none or one shock is permanent, the transformation recovers the correct representation. When both shocks are permanent, the representation is the first difference of (16). Instead of solving for a model that is possibly non-stationary, we solve for a model that is possibly over-differenced.

In this model, we estimate $(\alpha, \rho_z, \sigma_z, \rho_q, \sigma_q)$. The relative persistence and variability of shocks is important for the estimates. We fix $\sigma_z = 0.1$ and let σ_q take values (0.025,0.05,0.15). The persistence of the shocks to technology and hours is described by the vectors (0.95,0.99,1) and (0.5,0.8,0.9,0.975) respectively so that in our exercise technology shocks are generally more persistent than shocks to hours. To preserve space, we report only selected results in Table 4 and Figures 5-7 ($\rho_q = 0.8$,$\rho_z = 0.95$) and provide only a concise summary of the results. Additional results are available upon request.

In short, combinations (QD,$\Delta^{\rho_z}\widehat{m}_t$) and (FD,$\Delta\widehat{m}_t$) perform well while other estimators have significantly worse performance. Using HP filter to remove the trend as in (HP,$\widehat{m}_t$) or (HP,$\widetilde{m}_t$) continues to induce very strong biases in all estimates. The combination (FD,$\Delta^1 m_t$) performs well when technology shocks have a unit root but its performance quickly deteriorates as ρ_z moves away from one. Linear detrending in (LT,$\widehat{m}_t$) can perform relatively well when shocks to stationary hours are large relative to technology shocks. That is, as shocks to hours explain a larger fraction of variation in variables, identification of structural parameters improves as one can rely on variation in stationary, non-persistent structural shocks. For example, in the case of $\rho_z = 0.95$, as σ_q increases from 0.025 to 0.15 the mean estimate of α falls from 0.4520 to 0.3813 so that the bias decreases from 0.1220 to 0.0513. The reduction in the bias is even more dramatic when ρ_z is closer to one. The bias in other estimates exhibits a similar pattern. However, the biases become pronounced again when shocks to hours become more persistent, i.e., ρ_q increases towards one. The pair (FD,$\Delta^1 m_t$) dominates (HP,$\widehat{m}_t$) or (HP,$\widetilde{m}_t$), either of which continues to exhibit strong biases in the estimates.

In some cases (e.g., (HP,$\widetilde{m}_t$)) the relative size of the shocks is reversed in the estimates—that is, $\widehat{\sigma}_q > \widehat{\sigma}_z$ while $\sigma_q < \sigma_z$—so that the researcher may be tempted to conclude that shocks to hours have larger volatility than shocks to technology while the opposite is true.

5 Concluding Remarks

This paper has several substantive findings. First, the paper identifies Problems (DD) and (MTS) and shows that the consequences of these two problems can be devastating for the estimates of structural parameters in DSGE models. Specifically, the paper demonstrates that Problems (DD) and (MTS) can lead to distorted estimates, spurious estimates of propagation/amplification mechanisms (both external and internal), poor identification of structural parameters (especially parameters identified from low frequency variation). Importantly, both Problem (DD) and Problem (MTS) are empirically relevant and often arise in applied work.

Second, the paper proposes a robust approach to address Problems (DD) and (MTS) simultaneously. Specifically the paper shows that quasi-differencing or similar approaches (e.g., first differencing) tackle both problems by applying the same transformation to the data and model variables and using the fact that this transformation yields stationary series for all parameters values that can describe persistence of forcing variables in the model. Simulations show that quasi-differencing outperforms popular alternatives not only in terms of having smaller bias and smaller dispersion of the estimates but also in terms of providing $\sqrt{T}$ consistent inference even for parameters governing persistence of exogenous shocks. Although the paper illustrates the working of the quasi-differencing estimator on specific simple examples, the paper also shows that the framework can be easily generalized to more complex settings.

References

Altig, D., Christiano, L., Eichenbaum, M. and Linde, J. 2004, Firm-Specific Capital, Nominal Rigidities, and the Business Cycle, FRB Cleveland WP 2004-16.

Altug, S. 1989, Time-to-Build and Aggregate Fluctuations: Some New Evidence, *International Economic Review* **30**, 889–920.

Bouakez, H., Cardia, M. and Ruge-Murcia, F. J. 2005, Habit Formation and the Persistence of Monetary Shocks, *Jornal of Monetary Economics* **52**, 1073–1088.

Burnside, C., Eichenbaum, M. and Rebelo, S. 1993, Labor Hoarding and the Business Cycle, *Journal of Political Economy* **101**, 245–273.

Christiano, L. and den Haan, W. 1996, Small-Sample Properties of GMM for Business Cycle Analysis, *Jornal of Business and Economic Statistics* **14**, 309–327.

Christiano, L. and Eichenbaum, M. 1992, Current Real-Business-Cycle Theories and Aggregate Labor-Market Fluctuations, *American Economic Review* **82**, 430–450.

Christiano, L., Eihenbaum, M. and Evans, C. 2005, Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy, *Quarterly Journal of Economics* **113**, 1–45.

Clarida, R., Gali, J. and Gertler, M. 2000, Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory, *Quarterly Journal of Economics* **115**, 147–180.

Cogley, T. 2001, Estimating and Testing Rational Expectations Models When the Trend Specification is Uncertain, *Jornal of Economic Dynamics and Control* **25**, 1485–1525.

Cogley, T. and Nason, J. 1995, Effects of the Hodrick-Prescott Filter on Trend and Difference Stationary Time Series: Implications for Business Cycle research, *Jornal of Economic Dynamics and Control* **19**, 253–278.

Del Negro, M., Schorfheide, F., Smets, F. and Wouters, R. 2007, On the Fit and Forecasting Performance of New Keynesian Models, *Jornal of Business and Economic Statistics* **25**, 123–143.

Dib, A. 2003, An Estimated Canadian DSGE Model with Nominal and Real Rigidities, *Canadian Journal of Economics* **36**, 949–972.

Doorn, D. 2006, Consequences of Hodrick-Prescott Filtering for Parameter Estimation in a Structural Model of Inventory, *Applied Economics* **38**, 1863–1875.

Edge, R., Kiley, M. and Laforte, J. P. 2005, An Estimated DSGE Model of the U.S. Economy, mimeo.

Faia, E. 2007, Finance and International Business Cycles, *Jornal of Monetary Economics* **54**, 1018–1034.

Fuhrer, J. 1997, The (Un)Importance of Forward-Looking Behavior in Price Specifications, *Journal of Money, Credit, and Banking* **29**, 338–350.

Fuhrer, J. and Rudebusch, G. 2004, Estimating the Euler Equation for Output, *Jornal of Monetary Economics* **51**, 1133–1153.

Gorodnichenko, Y. and Ng, S. 2007, A Simple root-T Consistent and Asymptotically Normal Estimator for the Largest Autoregressive Root, mimeo.

Harvey, A. and Jaeger, A. n.d., Detrending, Stylized Factors, and the Business Cycle, *Jornal of Applied Econometrics* **8**, 231–247.

Hayashi, F. 2000, *Econometrics*, Princeton University Press.

Ireland, P. 2001, Sticky-price models of the business cycle: Specification and stability, *Jornal of Monetary Economics* **47**, 3–18.

Ireland, P. 2004a, A Method for Taking Models to the Data, *Jornal of Economic Dynamics and Control* **28**, 1025–1226.

Ireland, P. 2004b, Technology Shocks in the New Keynesian Model, *Review of Economics and Statistics* **86**, 923936.

Kim, J. 2000, Constructing and Estimating a Realistic Optimizing Model of Monetary Policy, *Jornal of Monetary Economics* **45**, 329–359.

King, R. and Rebelo, S. 1993, Low Frequency Filtering and Real Business Cycles, *Jornal of Economic Dynamics and Control* **17**, 207–231.

Kydland, F. and Prescott, E. 1982, Time to Build and Aggregate Fluctuations, *Econometrica* **50**, 1345–1370.

McGrattan, E., Rogerson, R. and Wright, R. 1997, An Equilibrium Model of the Business Cycle with Household Production and Fiscal Policy, *International Economic Review* **38**, 267–290.

Nelson, C. and Kang, H. 1981, Spurious Periodicity in Inappropriately Detrended time Series, *Econometrica* **49**, 741–751.

Nelson, C. and Plosser, C. 1982, Trends and Random Walks in Macroeconmic Time Series: Some Evidence and Implications, *Jornal of Monetary Economics* **10**, 139–162.

Ruge-Murcia, F. 2005, Methods to Estimate Dynamic Stochastic General Equilibrium Models, *Jornal of Economic Dynamics and Control*.

Sbordone, A. 2006, U.S. Wage and Price Dynamics: A Limited Information Approach.

Singleton, K. J. 1988, Econometric Issues in the Analysis of Equilibrium Business Cycle Models, *Jornal of Monetary Economics* **21**, 361–368.

Smets, F. and Wouters, R. 2003, An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area, *Journal of the European Economics Assocation* **1:5**, 1123–1175.

Smets, F. and Wouters, R. 2007, Shocks and Frictions in US Business Cycles: A Baeysian DSGE Approach, *American Economic Review* **97**, 586–606.

Table 1. Summary of selected works.

Paper	Technology process	System vs. single equation	Filter	Estimation method
Ireland (2001)	Stationary AR	system	Linear trend	MLE
Del Negro, Schorfheide, Smets, and Wouters (2004)	Unit root with serial correlation in growth rates	system	First difference	Bayesian
Bouakez, Cardia, and J. Ruge-Murcia (2005)	Stationary AR	system	Linear trend	MLE
Faia (2007)	Stationary AR	system	HP	calibration
Clarida, Gali, and Gertler (2000)	Stationary AR	equation	HP and deviation from CBO measure of potential output	GMM
Christiano, Eihenbaum, and Evans (2005)	Not specified	system	VAR	GMM
Dib (2003)	Stationary AR	system	Linear trend	MLE
Smets and Wouters (2007)	Stationary AR	system	First difference	Bayesian
Smets and Wouters (2003)	Stationary AR	system	HP	Bayesian
Kim (2000)	Stationary AR	system	Linear trend	MLE
McGrattan, Rogerson, and Wright (1997)	Stationary AR	system	Linear trend and HP	MLE
Altug (1989)	Unit root	system	First differences	MLE in frequency domain
Fuhrer and Rudebusch (2004)	Not specified	equation	HP, one-sided BP, CBO, linear trend with breaks, quadratic deterministic trend	MLE, GMM
Fuhrer (1997)	Not specified	equation	HP, linear trend, quadratic trend	GMM
Kydland and Prescott (1982)	Permanent and transitory components	system	HP	calibration
Altig, Christiano, Eichenbaum, and Linde (2004)	Unit root with serial correlation in growth rates	system	First difference	GMM
Ireland (2004)	Unit root	system	Stationary variables and growth rates of nonstationary variables	MLE
Christiano and Eichenbaum (1992)	Unit root	system	HP	GMM
Burnside, Eichenbaum and Rebelo (1993)	Stationary AR	system	HP	GMM

Table 2. Basic one-shock model.

DGP	ρ_z		$(\mathrm{LT}, \hat{m}_t)$ (1)	$(\mathrm{HP}, \hat{m}_t)$ (2)	$(\mathrm{HP}, \tilde{m}_t)$ (3)	$(\mathrm{QD}, \Delta^{\rho}\hat{m}_t)$ (4)	$(\mathrm{FD}, \Delta^{1}\hat{m}_t)$ (5)	$(\mathrm{FD}, \Delta\hat{m}_t)$ (6)
			$\hat{\alpha}$: capital intensity, $\alpha = 0.33$					
DT	0.50	mean	0.3467	0.4201	0.9900	0.3435	0.3880	0.3330
		sd	0.0264	0.0066	0.0000	0.0265	0.0162	0.0171
		median	0.3462	0.4207	0.9900	0.3438	0.3866	0.3293
DT	0.95	mean	0.4554	0.6245	0.9900	0.3260	0.4227	0.3552
		sd	0.0996	0.0090	0.0000	0.0414	0.0596	0.0819
		median	0.4500	0.6250	0.9900	0.3246	0.4136	0.3309
DT	0.99	mean	0.7998	0.7430	0.9900	0.3241	0.3829	0.3607
		sd	0.1931	0.0133	0.0000	0.0407	0.0850	0.0935
		median	0.8835	0.7434	0.9900	0.3228	0.3591	0.3307
ST	1.00	mean	0.9525	0.7804	0.9900	0.3178	0.3545	0.3499
		sd	0.1370	0.0135	0.0000	0.0362	0.0886	0.0883
		median	0.9747	0.7804	0.9900	0.3210	0.3284	0.3285
			$\hat{\sigma}$: st.dev. of shocks to technology, $\sigma = 0.1$					
DT	0.50	mean	0.1048	0.1237	0.0037	0.1033	0.1840	0.1009
		sd	0.0091	0.0064	0.0003	0.0086	0.0087	0.0101
		median	0.1039	0.1226	0.0037	0.1024	0.1836	0.0990
DT	0.95	mean	0.1076	0.1757	0.0047	0.0985	0.1380	0.1095
		sd	0.0178	0.0121	0.0005	0.0124	0.0209	0.0309
		median	0.1039	0.1753	0.0047	0.0977	0.1338	0.1004
DT	0.99	mean	0.3377	0.2413	0.0043	0.0990	0.1175	0.1101
		sd	0.1660	0.0211	0.0005	0.0106	0.0259	0.0286
		median	0.3429	0.2409	0.0043	0.0980	0.1094	0.1002
ST	1.00	mean	3.3506	0.2749	0.0041	0.0976	0.1077	0.1062
		sd	1.7781	0.0255	0.0004	0.0088	0.0239	0.0237
		median	2.8516	0.2738	0.0041	0.0970	0.0991	0.0986
			$\hat{\rho}_z$: persistence of shocks to technology					
DT	0.50	mean	0.4611	0.2441	1.0000	0.4809	1.0000	0.4942
		sd	0.0651	0.0411		0.0549		0.0810
		median	0.4650	0.2455		0.4860		0.4947
DT	0.95	mean	0.9270	0.5319	1.0000	0.9449	1.0000	0.9508
		sd	0.0266	0.0329		0.0140		0.0127
		median	0.9350	0.5329		0.9464		0.9498
DT	0.99	mean	0.9120	0.5049	1.0000	0.9899	1.0000	0.9897
		sd	0.0621	0.0281		0.0045		0.0055
		median	0.9218	0.5070		0.9906		0.9896
ST	1	mean	0.6932	0.4871	1.0000	0.9993	1.0000	0.9981
		sd	0.0910	0.0257		0.0020		0.0026
		median	0.6724	0.4890		1.0000		0.9991

Note: This table presents summary statistics for estimates of $\alpha = 0.33$, $\sigma = 0.1$, and $\rho_z = (0.5, 0.95, 0.99, 1)$. The number of simulations is 500. sample size is T=300. In the top-row label (XX,YY), XX denotes the method of detrending and YY indicates the model concept of the observed variables. LT is linear detrending, HP is Hodrick-Prescott filter, FD is first differencing, QD is quasi differencing. Δ^1 denotes the restriction $\rho_z = 1$ when the model is solved in first differences. $\Delta^1 = 1 - \rho_z L$ denotes quasi differencing.

Table 3. Basic one-shock model with habit formation in consumption.

DGP	ρ_z		$(\text{LT},\hat{m}_t)$ (1)	$(\text{HP},\hat{m}_t)$ (2)	$(\text{HP},\tilde{m}_t)$ (3)	$(\text{QD},\Delta^{\rho}\hat{m}_t)$ (4)	$(\text{FD},\Delta^{1}\hat{m}_t)$ (5)	$(\text{FD},\Delta\hat{m}_t)$ (6)
			$\hat{\alpha}$: capital intensity, $\alpha = 0.33$					
DT	0.95	mean	0.3991	0.7916	0.9647	0.3280	0.4399	0.3336
		sd	0.0599	0.0441	0.0202	0.0248	0.0196	0.0260
		median	0.3923	0.7980	0.9687	0.3262	0.4380	0.3336
DT	0.99	mean	0.6067	0.8224	0.9836	0.3356	0.3582	0.3394
		sd	0.1988	0.1216	0.0090	0.0252	0.0291	0.0295
		median	0.6159	0.8630	0.9879	0.3366	0.3633	0.3386
ST	1.00	mean	0.8061	0.8357	0.9864	0.3387	0.3386	0.3406
		sd	0.1831	0.1705	0.0079	0.0282	0.0362	0.0288
		median	0.8876	0.9008	0.9896	0.3365	0.3422	0.3358
			$\hat{\sigma}$: st.dev. of shocks to technology, $\sigma = 0.1$					
DT	0.95	mean	0.0881	0.1758	0.0066	0.0972	0.1171	0.1009
		sd	0.0132	0.0268	0.0023	0.0083	0.0095	0.0051
		median	0.0868	0.1808	0.0062	0.0967	0.1164	0.1008
DT	0.99	mean	0.0630	0.1271	0.0042	0.0984	0.1028	0.1011
		sd	0.0179	0.1266	0.0011	0.0055	0.0052	0.0048
		median	0.0608	0.0852	0.0038	0.0986	0.1024	0.1009
ST	1.00	mean	0.0640	0.0452	0.0038	0.0997	0.1003	0.1013
		sd	0.0330	0.0617	0.0010	0.0048	0.0046	0.0045
		median	0.0622	0.0368	0.0035	0.0996	0.1004	0.1010
			$\hat{\rho}_z$: persistence of shocks to technology					
DT	0.95	mean	0.9434	0.6450	1.0000	0.9453	1.0000	0.9485
		sd	0.0070	0.0670		0.0103		0.0088
		median	0.9449	0.6431		0.9460		0.9496
DT	0.99	mean	0.9870	0.8058	1.0000	0.9875	1.0000	0.9879
		sd	0.0040	0.1809		0.0055		0.0071
		median	0.9881	0.8565		0.9886		0.9891
ST	1	mean	0.9908	0.9439	1.0000	0.9980	1.0000	0.9962
		sd	0.0077	0.0705		0.0033		0.0052
		median	0.9926	0.9601		0.9996		0.9983
			$\hat{\phi}$: habit formation in consumption, $\phi = 0.8$					
DT	0.95	mean	0.8280	0.9262	0.8255	0.7981	0.7536	0.8032
		sd	0.0218	0.0757	0.2003	0.0091	0.0219	0.0216
		median	0.8275	0.9336	0.8952	0.7979	0.7545	0.8033
DT	0.99	mean	0.8386	0.7749	0.8518	0.8025	0.7960	0.8074
		sd	0.1763	0.4104	0.1799	0.0150	0.0148	0.0241
		median	0.8906	0.9362	0.9082	0.8047	0.7986	0.8059
ST	1	mean	0.2864	0.7555	0.8221	0.8042	0.8033	0.8115
		sd	0.3918	0.3366	0.2081	0.0140	0.0205	0.0219
		median	0.2892	0.9515	0.8899	0.8032	0.8070	0.8060

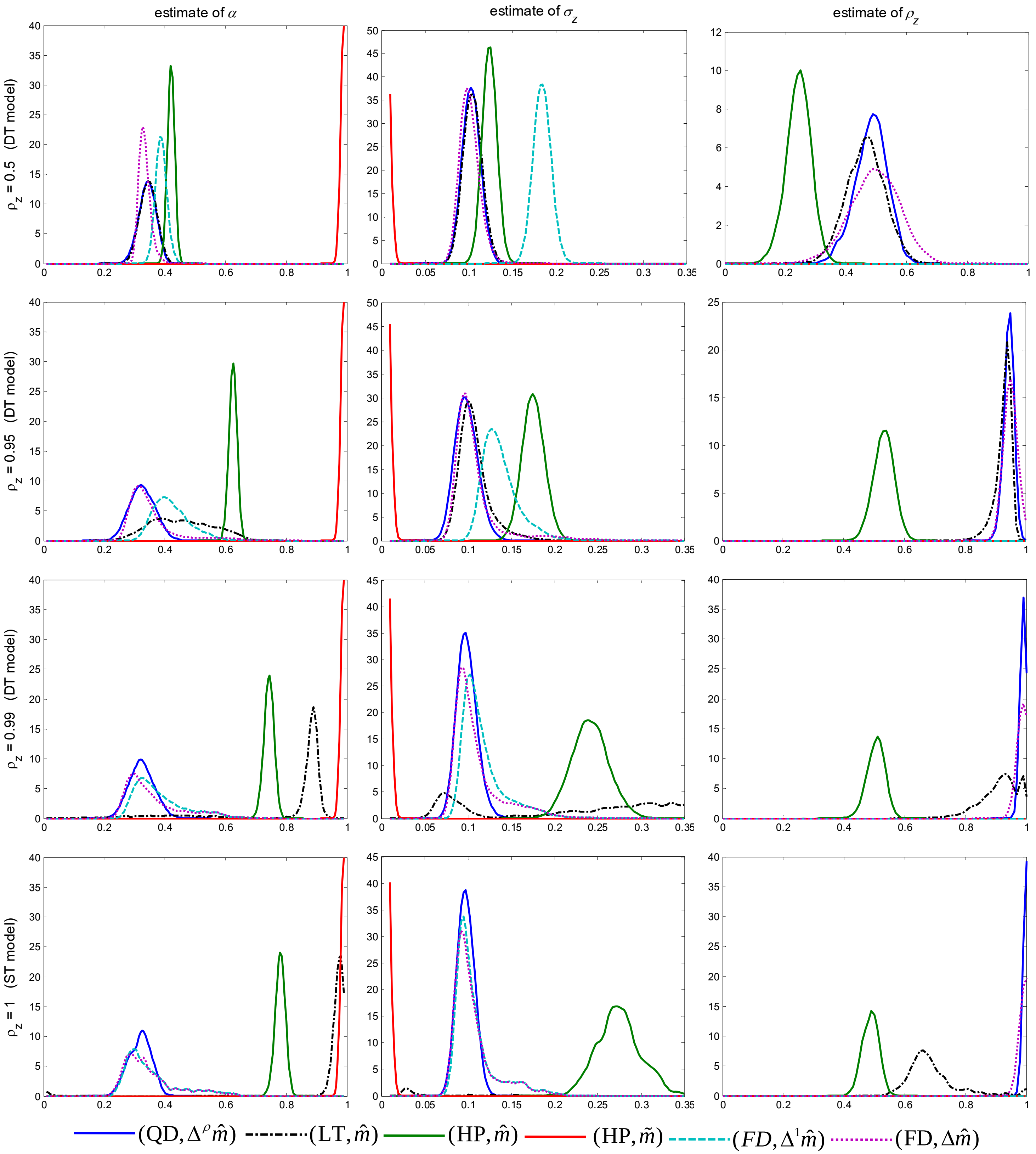
Note: This table presents summary statistics for estimates of $\alpha = 0.33$, $\sigma = 0.1$, and $\rho_z = (0.95, 0.99, 1)$. The number of simulations is 500. sample size is T=300. In the top-row label (XX,YY), XX denotes the method of detrending and YY indicates the model concept of the observed variables. LT is linear detrending, HP is Hodrick-Prescott filter, FD is first differencing, QD is quasi differencing. Δ^1 denotes the restriction $\rho_z = 1$ when the model is solved in first differences. $\Delta^1 = 1 - \rho_z L$ denotes quasi differencing.

Table 4. Two-shock model, estimate of α.

DGP	ρ_z		$(\text{LT},\hat{m}_t)$	$(\text{HP},\hat{m}_t)$	$(\text{HP},\tilde{m}_t)$	$(\text{QD},\Delta^{\rho}\hat{m}_t)$	$(\text{FD},\Delta^{1}\hat{m}_t)$	$(\text{FD},\Delta\hat{m}_t)$
			(1)	(2)	(3)	(4)	(5)	(6)
					σ_q = 0.025			
DT	0.95	mean	0.4520	0.6516	0.4487	0.3385	0.2911	0.3573
		sd	0.1054	0.0278	0.1392	0.0285	0.0453	0.0766
		median	0.4394	0.6510	0.4043	0.3310	0.2789	0.3326
DT	0.99	mean	0.7798	0.7384	0.8887	0.3444	0.3499	0.3667
		sd	0.1964	0.0323	0.1863	0.0356	0.0816	0.0891
		median	0.8552	0.7394	0.9840	0.3326	0.3217	0.3360
ST	1.00	mean	0.9197	0.7619	0.9497	0.3448	0.3622	0.3644
		sd	0.1207	0.0326	0.1208	0.0359	0.0863	0.0865
		median	0.9473	0.7626	0.9859	0.3330	0.3292	0.3312
					σ_q = 0.05			
DT	0.95	mean	0.4442	0.6323	0.4031	0.3354	0.2900	0.3491
		sd	0.1018	0.0393	0.0361	0.0194	0.0225	0.0556
		median	0.4315	0.6279	0.3974	0.3316	0.2856	0.3334
DT	0.99	mean	0.7489	0.6914	0.4546	0.3376	0.3269	0.3507
		sd	0.2020	0.0425	0.0512	0.0207	0.0356	0.0594
		median	0.8362	0.6896	0.4450	0.3332	0.3178	0.3322
ST	1.00	mean	0.8794	0.7065	0.4725	0.3422	0.3432	0.3561
		sd	0.1463	0.0436	0.0606	0.0246	0.0407	0.0578
		median	0.9292	0.7047	0.4602	0.3374	0.3340	0.3377
					σ_q = 0.15			
DT	0.95	mean	0.3813	0.5722	0.5378	0.3354	0.3225	0.3417
		sd	0.0581	0.0471	0.0434	0.0123	0.0115	0.0310
		median	0.3735	0.5679	0.5362	0.3331	0.3208	0.3311
DT	0.99	mean	0.4182	0.5878	0.5481	0.3383	0.3316	0.3455
		sd	0.1270	0.0446	0.0435	0.0125	0.0120	0.0298
		median	0.3761	0.5846	0.5453	0.3363	0.3302	0.3351
ST	1.00	mean	0.4433	0.5900	0.5512	0.3399	0.3344	0.3469
		sd	0.1640	0.0445	0.0431	0.0127	0.0120	0.0291
		median	0.3759	0.5866	0.5483	0.3380	0.3329	0.3371

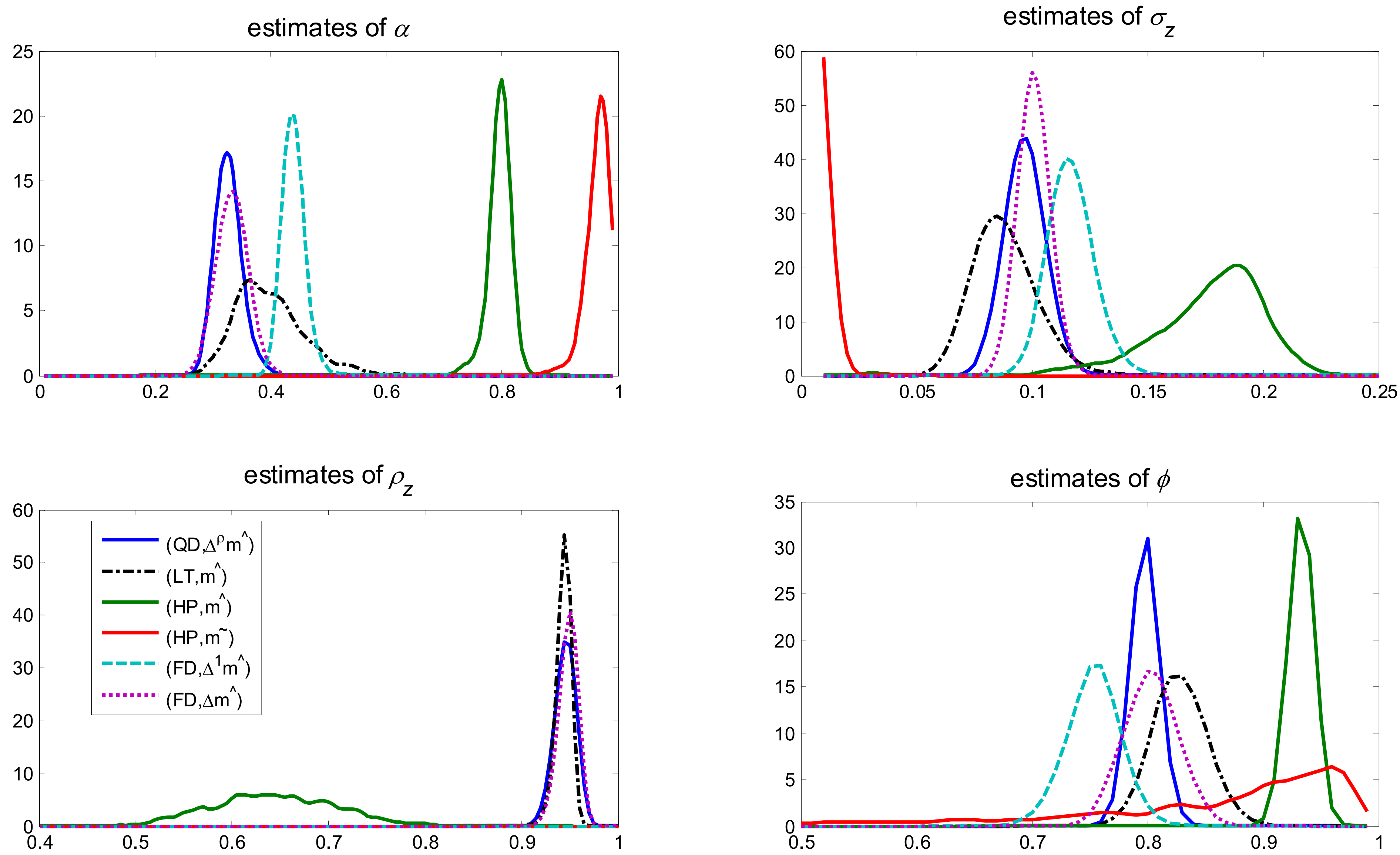
Note: Other parameters are fixed at $\alpha = 0.33, \rho_q = 0.8, \sigma_z = 0.1$. 1000 simulations. Sample size T=300. In the top-row label (XX,YY), XX denotes the method of detrending and YY indicates the model concept of the observed variables. LT is linear detrending, HP is Hodrick-Prescott filter, FD is first differencing, QD is quasi differencing. Δ^1 denotes the restriction $\rho_z = 1$ when the model is solved in first differences. $\Delta^{\rho} = 1 - \rho L$ denotes quasi differencing.

Figure 1. Kernel density of estimates, baseline model.



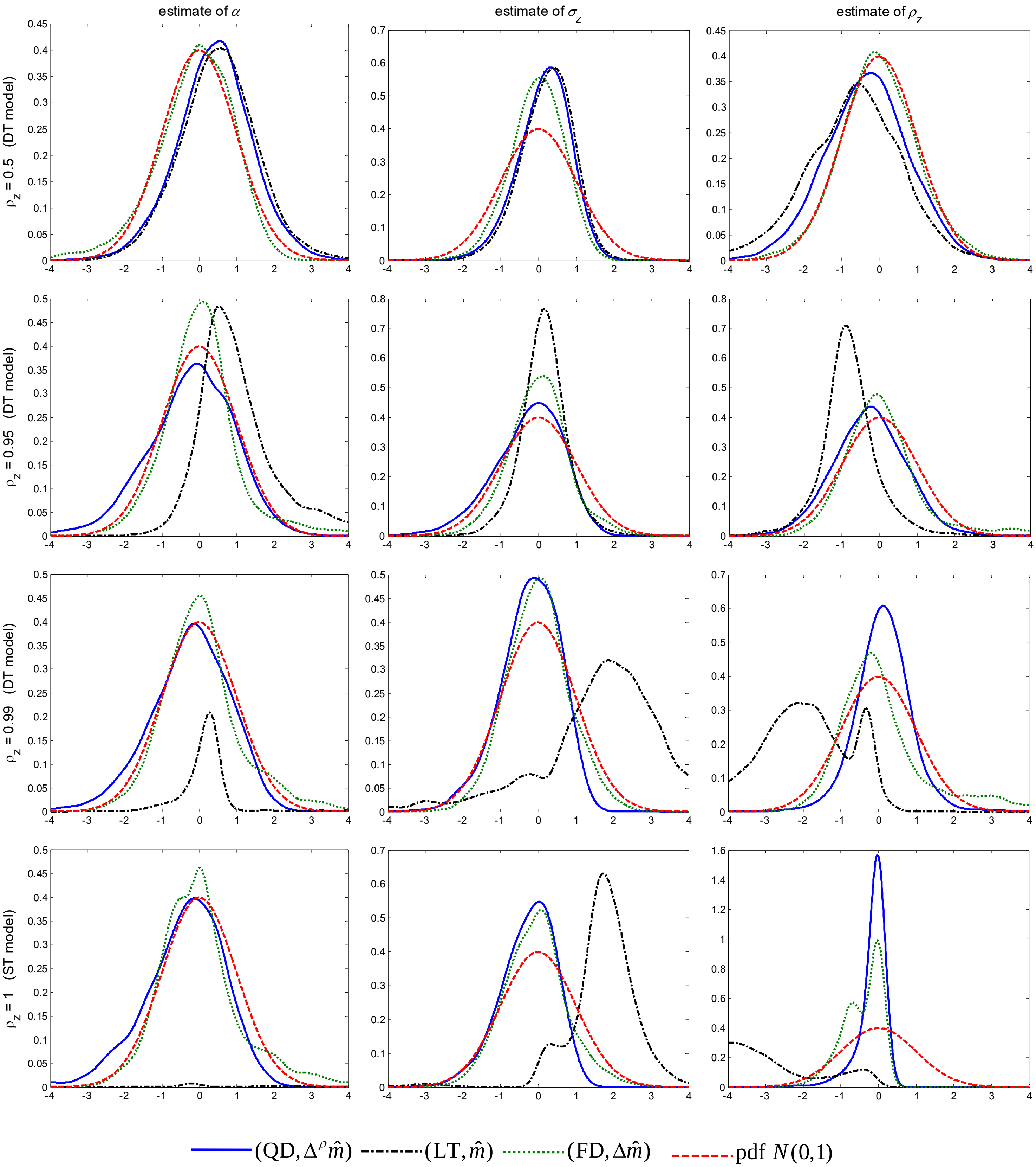
Note: This figure plots kernel density of $(\hat{\alpha}, \hat{\rho}, \hat{\sigma})$ generated in 1000 simulations. Bandwidth is 0.01. For the case presented in this figure the true values are $\alpha = 0.33, \sigma = 0.1$. True value of ρ_z is indicated on the left of the figure. In the legend (XX,YY), XX denotes the method of detrending and YY indicates the model concept of the observed variables.

Figure 2. Kernel density of estimates for the model with habit formation, $\rho_z = 0.95$.



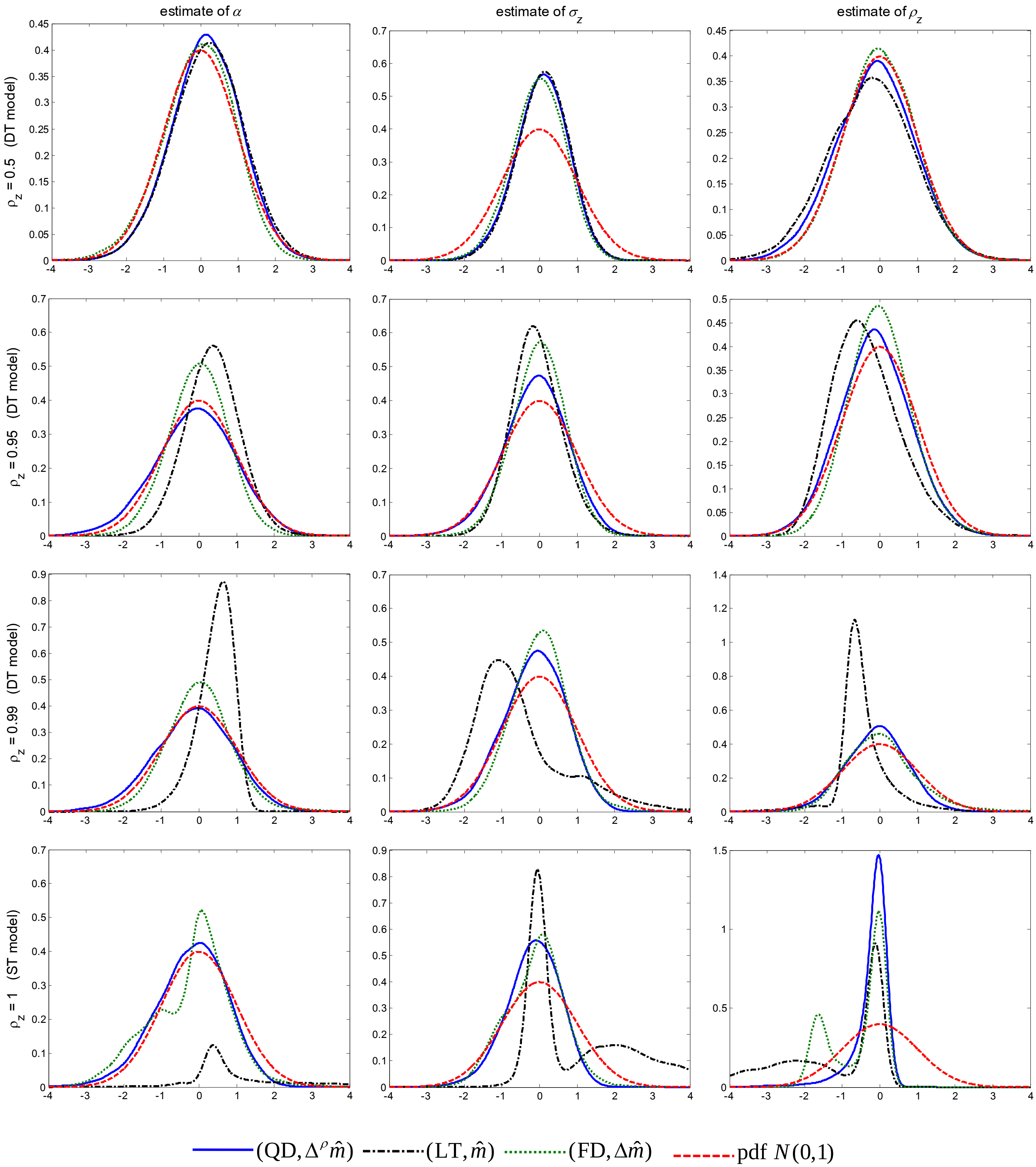
Note: This figure plots kernel density of $(\hat{\alpha}, \hat{\rho}, \hat{\sigma}, \hat{\phi})$ generated in 1000 simulations. Bandwidth is 0.01. For the case presented in this figure the true values are $\alpha = 0.33, \rho_z = 0.95, \phi = 0.8, \sigma = 0.1$. In the legend (XX,YY), XX denotes the method of detrending and YY indicates the model concept of the observed variables.

Figure 3. Kernel density for t-statistic, baseline model, T=300.



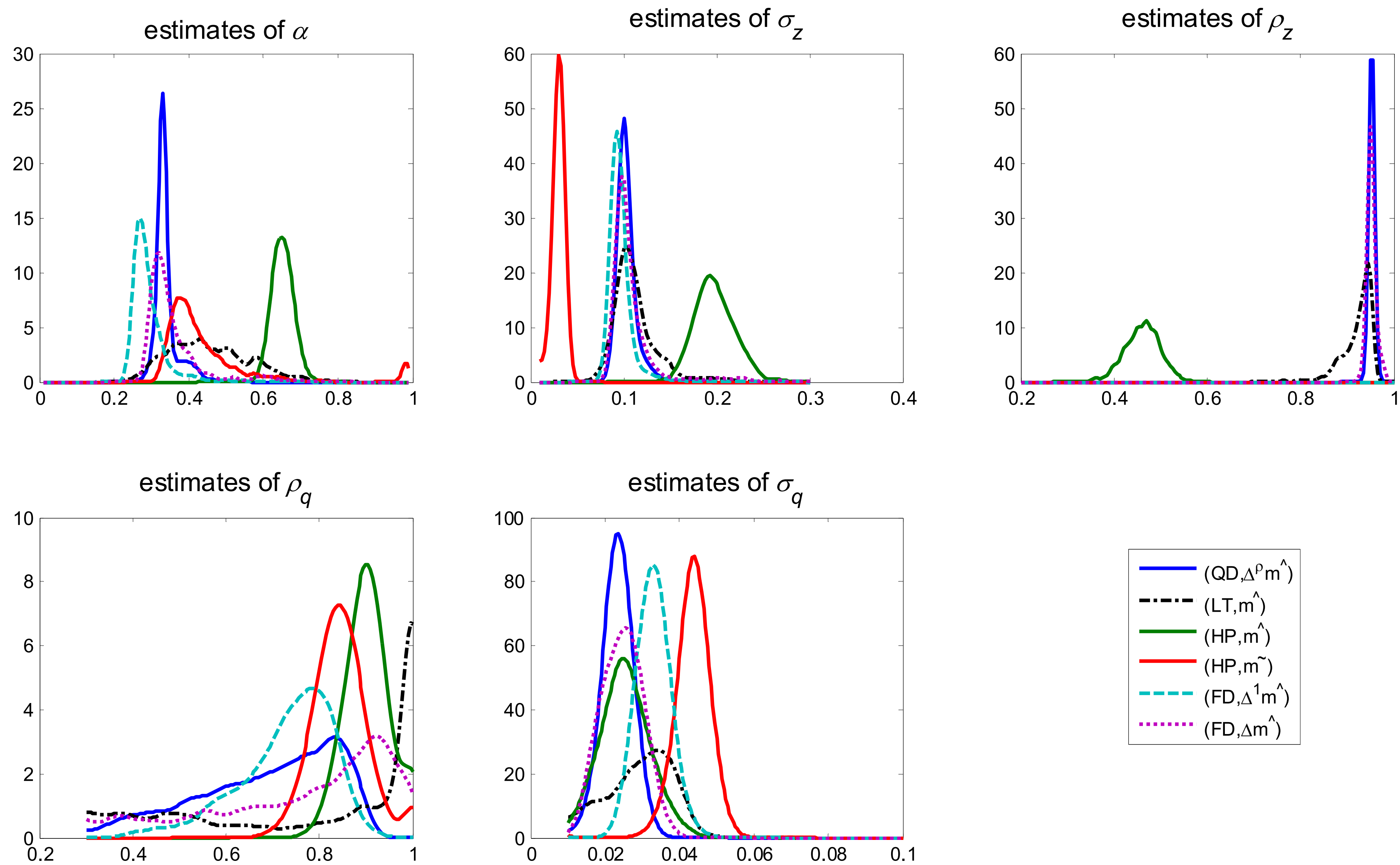
Note: This figure plots kernel density of t-statistic for $(\hat{\alpha}, \hat{\rho}, \hat{\sigma})$ generated in 1000 simulations. Bandwidth is 0.01. In the legend (XX,YY), XX denotes the method of detrending and YY indicates the model concept of the observed variables.

Figure 4. Kernel density for t-statistic, baseline model, T=2000.



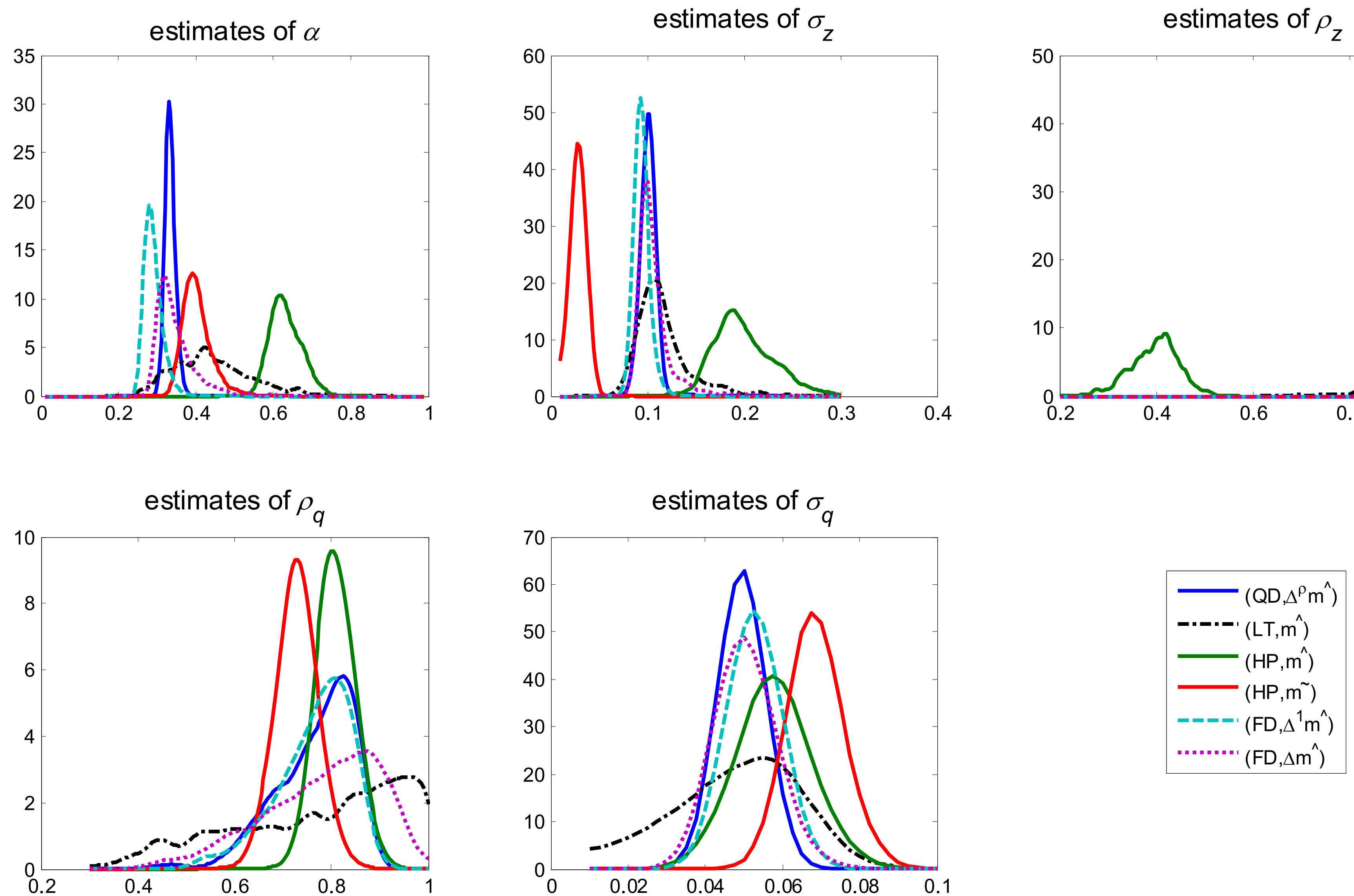
Note: This figure plots kernel density of t-statistic for $(\hat{\alpha}, \hat{\rho}, \hat{\sigma})$ generated in 1000 simulations. Bandwidth is 0.01. In the legend (XX,YY), XX denotes the method of detrending and YY indicates the model concept of the observed variables.

Figure 5. Kernel density of estimates for the model with shocks to hours and technology, $\rho_z = 0.95$, $\sigma_q = 0.025$.



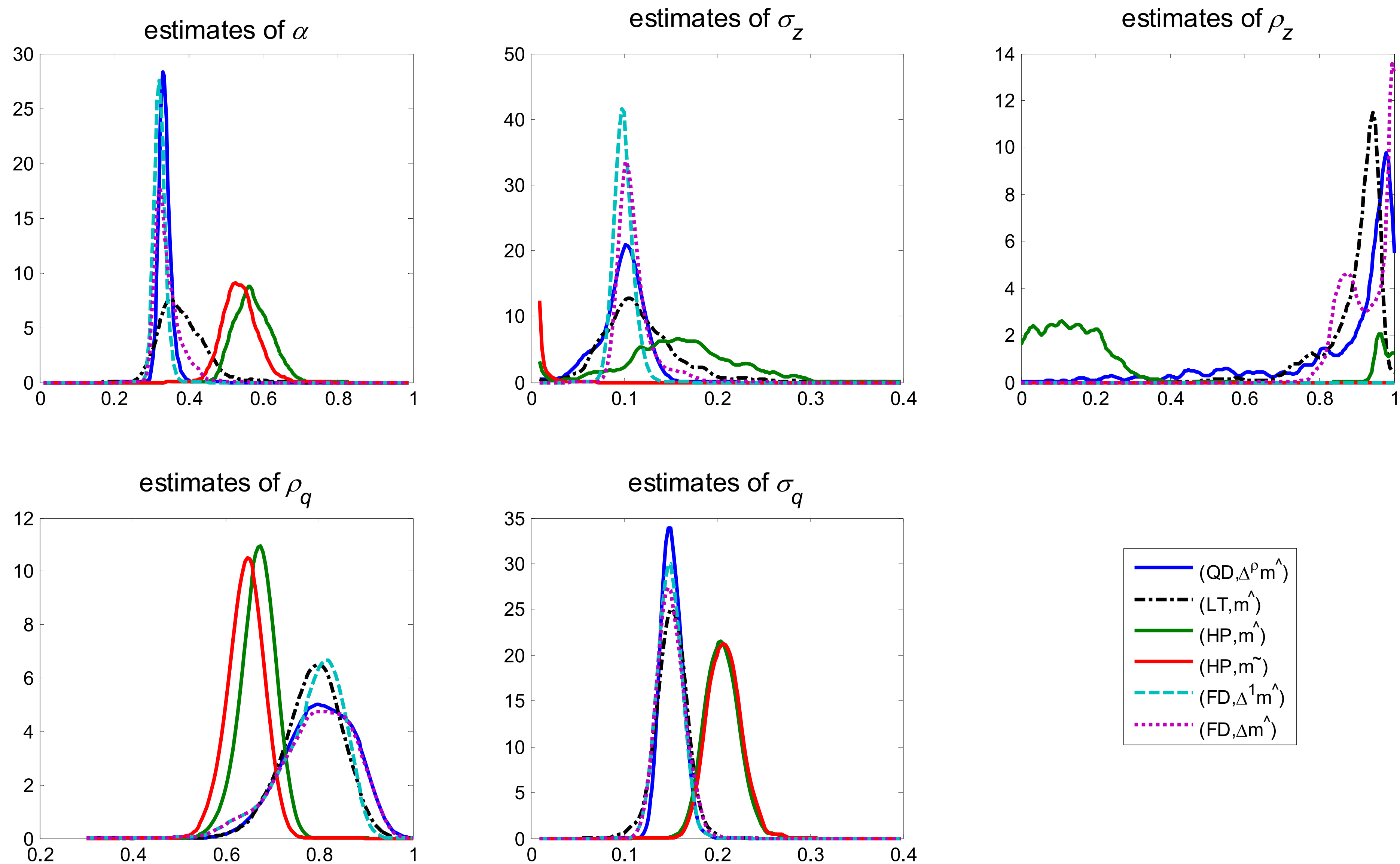
Note: This figure plots kernel density of $(\hat{\alpha}, \hat{\rho}_z, \hat{\sigma}_z, \hat{\rho}_q, \hat{\sigma}_q)$ generated in 1000 simulations. Bandwidth is 0.01. For the case presented in this figure the true values are $\alpha = 0.33, \rho_z = 0.95, \sigma_z = 0.1, \rho_q = 0.8, \sigma_z = 0.025$. In the legend (XX,YY), XX denotes the method of detrending and YY indicates the model concept of the observed variables. See text for further details.

Figure 6. Kernel density of estimates for the model with shocks to hours and technology, $\rho_z = 0.95$, $\sigma_q = 0.05$.



Note: This figure plots kernel density of $(\hat{\alpha}, \hat{\rho}_z, \hat{\sigma}_z, \hat{\rho}_q, \hat{\sigma}_q)$ generated in 1000 simulations. Bandwidth is 0.01. For the case presented in this figure the true values are $\alpha = 0.33, \rho_z = 0.95, \sigma_z = 0.1, \rho_q = 0.8, \sigma_z = 0.05$. In the legend (XX,YY), XX denotes the method of detrending and YY indicates the model concept of the observed variables. See text for further details.

Figure 7. Kernel density of estimates for the model with shocks to hours and technology, $\rho_z = 0.95$, $\sigma_q = 0.15$.



Note: This figure plots kernel density of $(\hat{\alpha}, \hat{\rho}_z, \hat{\sigma}_z, \hat{\rho}_q, \hat{\sigma}_q)$ generated in 1000 simulations. Bandwidth is 0.01. For the case presented in this figure the true values are $\alpha = 0.33, \rho_z = 0.95, \sigma_z = 0.1, \rho_q = 0.8, \sigma_z = 0.15$. In the legend (XX,YY), XX denotes the method of detrending and YY indicates the model concept of the observed variables. See text for further details.