ESTIMATION OF DSGE MODELS WHEN THE DATA ARE PERSISTENT

Yuriy Gorodnichenko*   Serena Ng†

May 5, 2008

Abstract

An active area of research in macroeconomics is to take DSGE models to the data. These models are often solved and estimated 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. Estimates of parameters governing transmission mechanisms can be severely biased. We then consider four estimators that are robust to whether shocks in the model are assumed to be permanent or transitory and do not require the researcher to take a stand on the dynamic properties of the data. Simulations show that when the shocks are not persistent, the proposed estimators are as precise as estimators that correctly impose the stationarity assumption. But when the shocks are highly persistent yet stationary, the proposed estimators are much more precise. These properties hold even when there are multiple persistent shocks.

*Department of Economics, UC Berkeley, Berkeley, CA 94720 Email: ygorodni@econ.berkeley.edu.

†Department of Economics, Columbia University, 420 West 118 St, MC 3308, New York, NY 10027 Email: Serena.Ng@columbia.edu.

This paper was presented at the University of Michigan, the 2007 NBER Summer Institute, and the New York Area Macro Conference. We thank Marc Giannone and Tim Cogley for many helpful discussions and the seminar participants for many helpful comments. The second author acknowledges financial support from the National Science Foundation (SES 0549978).

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.

As is well known, economic data are highly persistent and possibly non-stationary. It is common practice to allow shocks in DSGE models to have persistent effects. 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 or balanced growth path. 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. 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. 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	→	Data Detrending	→	Estimation
Problems:		MTS		DD	

Problem (DD) is concerned with how the observed data are filtered. The issue can arise when the detrended data do not have the same properties as those implicit in the model. For example, the model may specify the shock process to be 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. This issue can arise if, for example, the model assumes that technology is trend stationary and the data are linearly detrended accordingly, but the data contain stochastic trends. In consequence, the detrended data will still be non-stationary. As is well known, spurious detrending can invalidate classical inference. The problem that confronts researchers is that in finite samples, it is very difficult to ascertain whether the data are stationary or not. Yet, estimation of DSGE models typically require that the researcher takes a stand on trend specification both for the model and the data. Discrepancies between the properties of the cyclical component in the model and in the data can induce bias estimates of the structural parameters.

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. Stochastic trends are assumed in some studies and the first differenced data are then used in estimation. But since the seminal work of Nelson and Plosser (1982), there has been ongoing debate whether trend or difference stationary is a better characterization of macroeconomic variables. 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 non-linear estimation of DSGE models. This paper sheds some light on this issue.

In an early contribution, Nelson and Kang (1981) showed that linear detrending a unit root process can generate spurious cycles. Subsequent studies found that improper filtering can alter the persistence and the volatility of the series as discussed in Cogley and Nason (1995), induce spurious correlations in the filtered data as found in Harvey and Jaeger (1993), change error structure as shown in Singleton (1988)), and distort inference as illustrated in Christiano and den Haan (1996). However, much of this literature is focused on single equation analysis. Fukac and Pagan (2006) consider how the treatment of trends might affect estimation of DSGE models, but the analysis is also confined to a single equation framework. As DSGE models consists of a system of equations, misspecification in one equation can affect estimates in other equations even though the estimates should be more efficient if the model is correctly specified. King and Rebelo (1993) simulate an RBC model and show that the HP (Hodrick-Prescott) filtered data are qualitatively different from the raw data. Gregory and Smith (1996) use a *calibrated* business cycle model to see what type of trend can produce a cyclical component in the data similar to the cyclical component in the model. Although these authors do not estimate the model on filtered data, they hint that the estimates of the structural parameters can be adversely affected by filtering.

We use a basic stochastic growth model to illustrate the problems under consideration. When the trends assumed for the model agree with the trends present in the data *and* the same filter is applied to the model and the data, the estimated parameters are mean and median unbiased. Otherwise, the estimates can deviate significantly from the true values. Estimates of parameters governing the propagation and amplification mechanisms in the model can be greatly distorted or poorly identified.

We propose a robust strategy to handle uncertainty as to whether the data are trend or difference stationary. The proposed approach consists of applying the same transformation (filter) to both the data and the model variables, and making sure that the transformed series are stationary when evaluated at the true parameter vector. We illustrate this approach with four transformations: quasi-differencing, (unconstrained) first differencing, a hybrid quasi/first differenced filter, and the HP filter. The approach is shown to be effective even when there are multiple shocks, a subset of which may be permanent. Although our analysis is motivated as classical estimator, it can be adapted into a Bayesian framework.

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. We then illustrate how the two problems can yield misleading inference about the propagation mechanisms and extend the analysis to multiple structural shocks. The robust estimators are considered in Sections 4 and 5. Results for a two shock model is then presented.

2 An Example: Neoclassical Growth Model

We use the stochastic growth model to illustrate the problems under investigation. The general 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

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

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. We allow ρ_q and ρ_z to be on the unit circle. The first order conditions are:

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

If $\bar{g} = 0$ and $|\rho_z|, |\rho_q| < 1$, then under regularity conditions, a solution for the 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 $c_t - c_t^*$ is stationary; k_t^* and z_t^* are similarly defined. We assume labor L_t is stationary for all $|\rho_z| \leq 1$ and thus $l_t^* = 0$. Hereafter, we let $m_t = (c_t, k_t, l_t)$, and $m_t^* = (c_t^*, k_t^*, l_t^*)$. Note that m_t^* are model concepts. Where appropriate, we will drop Q_t to simplify the analysis.

2.1 The One Shock Model

To fix ideas, suppose for now that technology is the only shock in the system. Hence, Q_t is suppressed.

The DT Model When $|\rho_z| < 1$, let $c_t^* = k_t^* = \bar{g}t$, and thus $m_t^* = (\bar{g}t, \bar{g}t, 0)$. The detrended model variables are $\widehat{m}_t = m_t - m_t^*$. The log-linearized model in terms of $\widehat{m}_t$ is

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

where we have suppressed the constants terms and the matrices are defined as follows

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

Since a shock to technology has temporary effects, $\widehat{m}_t$ is stationary. We can compactly write (1) as

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

We will refer to (1) as the trend stationary (DT) representation of the model. The QZ decomposition or similar methods can be used to solve the system of expectation equations for the reduced form.

Denote this solution by

$$\widehat{m}_t = \Pi_{DT}\widehat{m}_{t-1} + B_{DT}u_t^z \tag{1}$$

with $u_t^z = \rho^z u_{t-1} + e_t^z$.

The ST Model When $|\rho_z| = 1$, let $c_t^* = k_t^* = z_t$ and thus $m_t^* = (z_t, z_t, 0)$. Let $\widetilde{m}_t = m_t - m_t^*$ denote the stationary model variables for the ST model. The log-linearized model expressed in terms of $\widetilde{m}_t$ is:

$$E_t \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & A_0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \widetilde{c}_{t+1} \\ \widetilde{k}_{t+1} \\ \widetilde{l}_{t+1} \end{bmatrix} = \begin{bmatrix} -1 & 0 & -\alpha \\ 1 & A_0 & 0 \\ A_1 & A_2 & \alpha - 1 \end{bmatrix} \begin{bmatrix} \widetilde{c}_t \\ \widetilde{k}_t \\ \widetilde{l}_t \end{bmatrix} + \begin{bmatrix} 0 & \alpha & 0 \\ 0 & 0 & 0 \\ 0 & A_4 & 0 \end{bmatrix} \begin{bmatrix} \widetilde{c}_{t-1} \\ \widetilde{k}_{t-1} \\ \widetilde{l}_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ -A_0 \\ 0 \end{bmatrix} E_t e_{t+1}^z + \begin{bmatrix} 1-\alpha \\ 0 \\ \alpha - 1 \end{bmatrix} e_t^z$$

or more compactly

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

We will refer to (2) as the stochastic trend (ST) representation of the model. The solution takes the form

$$\widetilde{m}_t = \Pi_{ST}\widetilde{m}_{t-1} + B_{ST}e_t^z. \tag{2}$$

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

$$\widetilde{m}_t = \widehat{m}_t - (u_t^z, u_t^z, 0).$$

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

The system of equations (1) and (2) both correspond to the same stochastic growth model. As one would expect, the rational expectations solution for variables in *levels* is the same irrespective of which model we solve. Although not expressed in the usual state-space representation form, the equations and system matrices make it clear that the models are distinguished only in what variables we analyze, i.e., $\widehat{m}_t$ for DT, and $\widetilde{m}_t$ for ST. The distinction between $\widetilde{m}_t$, m_t and $\widehat{m}_t$ is important because the former is stationary when $\rho_z = 1$ while $\widehat{m}_t$ is not. Importantly, it is these *normalized* variables that are linked through a measurement equation to the filtered data. As classical estimation assumes that the data are stationary, we should map $\widetilde{m}_t$ and not $\widehat{m}_t$ to the data when $\rho_z = 1$.

The Δ^1 DT Model When $\rho_z = 1$, the VAR in $\widehat{m}_t$ is non-stationary because u_t^z is non-stationary. However, first differencing gives

$$\Delta^1\widehat{m}_t = \Pi_{\Delta^1}\Delta^1\widehat{m}_{t-1} + B_{\Delta^1}e_t^z \tag{3}$$

is a stationary VAR. 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. Thus, when $\rho_z = 1$, we can work with $\Delta\widehat{m}_t$ or $\widetilde{m}_t$. One may also express the model in terms of stationary linear combinations of the non-stationary variables. For example, $c_t - y_t$ is stationary for all $|\rho_z| \leq 1$. Imposing cointegration and unit root restrictions as in the Δ^1DT model can be efficient when $\rho_z = 1$ but is not appropriate when ρ_z is close to but is not exactly equal to one.

2.2 Estimation

Suppose we observe the data for $d_t = (c_t, k_t, l_t)$. Let $d_t^c = d_t - d_t^\tau$ denote the data after the trend d_t^τ is removed. Three commonly used alternatives are:

- Linear Trend (LT): $d_t^c = d_t - d_t^\tau$ with $d_t^\tau = (\bar{g}t, \bar{g}t, 0)$;
- HP Trend (HP): $d_t^c = (HP(L)c_t, HP(L)k_t, 0)$
- First Difference (FD): $d_t^c = \Delta d_t$.[1]

Typically, linearly detrended and HP filtered 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 $\rho_z = 1$.

Likelihood based approaches (e.g. Fernandez-Villaverde and Rubio-Ramirez (2006) and Ireland (1997)), two-step minimum distance approach (e.g., Sbordone (2006)), as well as simulation estimation (e.g., Altig et al. (2004)) have been used to estimate DSGE models. Ruge-Murcia (2005) provides a review of these methods. We use a method of moments estimator that minimizes the distance between data moments and model-implied moments (e.g., Christiano and den Haan (1996), Christiano and Eichenbaum (1992)). The procedure can be summarized as follows:

1: Compute $\widehat{\Omega}^d(j) = cov(d_t^c)$, the covariance matrix of the filtered series at lag j.

2: Solve the rational expectations model for a guess of the structural parameters, Θ. Use (1), (2), or (3) to analytically compute $\Omega^m(j)$, the model implied autocovariances of m_t^c, which can be $\widehat{m}_t$, $\widetilde{m}_t$, or $\Delta\widehat{m}_t$.

[1] 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.

3: Let

$$\begin{aligned}
\widehat{\omega}^d &= (vech(\Omega^d(0))' \ vec(\Omega^d(1))' \ \ldots \ vec(\Omega^d(L))')' \\
\omega^m(\Theta) &= (vech(\Omega^m(0))' \ vec(\Omega^m(1))' \ \ldots \ vec(\Omega^m(L))')'.
\end{aligned}$$

Estimate the structural parameters as $\widehat{\Theta} = \text{argmin}_{\Theta} \left\| \widehat{\omega}^d - \omega^m(\Theta) \right\|$.

The two problems that motivated the present analysis are easily understood in the context of the covariance structure estimator. Problem (MTS) arises when m_t^* does not agree with d_t^τ, while Problem (DD) arises when the dynamic properties of d_t^c does not agree with the implied properties of m_t^c. The consequence is that $\omega^d - \omega(\Theta)$ will not be mean zero. To illustrate, consider the following combinations of model variables and data filtering techniques:

	True Model	Assumed Model	Variables	Filter	Problems
1.	DT	DT	$\widehat{m}_t$	LT	-
2.	ST	ST	$\Delta^1 \widehat{m}_t$	FD	-
3.	DT	DT	$\widehat{m}_t$	HP	(DD)
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, both (1) and (2) assume a model trend that 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 (3), the model assumes, and the data exhibit, deterministic trends, and there is no Problem (MTS). However, the HP filter applied to the data series produces cycles different from those that emerge from linear detrending. 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 and Problem (DD) does not arise. On the other hand, the assumed DT model 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).

How can these problems be resolved? Given that Problem (DD) arises when m_t^c of a given model solution is mapped to the data d_t^c, the problem can be circumvented by applying the same filtering technique to the model variables and the data series. Problem (MTS) concerns the appropriate choice of the model solution. The problem can be avoided if there 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 both Problems (DD) and (MTS), the researcher needs an approach that *i*) transforms the data and the model variables in the same way and *ii*) yields stationary series for all $|\rho_z| \leq 1$.

3 Four Robust Approaches

In this section, we consider four approaches that are robust to whether shocks are permanent or transitory. The key to robustness is to filter both the model variables and the observed data consistently so that filtered series are stationary and have the same properties. The starting point of all methods is the reduced form solution of the DT model:

$$\widehat{m}_t = \Pi_{DT}\widehat{m}_{t-1} + B_{DT}u_t^z$$

where $u_t^z = \rho^z u_{t-1} + e_t^z$. Under the DT model, ρ_z is a free parameter. Thus, adding deterministic terms back to the DT model yields a model that can potentially have both deterministic and stochastic trends.

3.1 The QD Estimator

Let $\Delta^{\rho_z} = 1 - \rho_z L$ be the quasi-differencing operator. Multiplying both sides of the reduced form solution by Δ^{ρ_z} and using $u_t^z = \rho_z u_{t-1}^z + e_t^z$ gives

$$\Delta^{\rho_z}\widehat{m}_t = \Pi_{DT}\Delta^{\rho_z}\widehat{m}_{t-1} + B_{DT}e_t^z \tag{4}$$

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 the quasi-differenced model 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, in contrast to $\widehat{m}_t$, the moments of $\Delta^{\rho}m_t$ are well defined for all $|\rho_z| \leq 1$. In addition, first differencing is just a special case of quasi-differencing with $\rho_z = 1$. If we partition $\Theta = (\Theta^-, \rho_z)$, the deep parameters can be estimated as follows:

Initialize ρ_z.

1: Quasi-difference the observed data with ρ_z to obtain $\Delta^{\rho_z}d_t^c = (c_t^c, k_t^c, l_t^c)$.[2] where

$$c_t^c = \Delta^{\rho_z}(c_t - \bar{g}t), \quad k_t^c = \Delta^{\rho_z}(k_t - \bar{g}t), \quad l_t^c = \Delta^{\rho_z}l_t. \tag{5}$$

[2] 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. Likewise, one can introduce structural breaks in a trend directly at this step.

2: Compute $\widehat{\Omega}^d_{\Delta^{\rho_z}}(j) = cov(\Delta^{\rho_z} d^c_t)$, the autocovariance matrix at lag j. Define

$$\Omega^d_{\Delta^{\rho_z}}(j) = \widehat{\Omega}^d_{\Delta^{\rho_z}}(j) - \widehat{\Omega}^d_{\Delta^{\rho_z}}(0)$$

and let $\widehat{\omega}^d_{\Delta^{\rho_z}} = (vec(\Omega^d_{\Delta^{\rho_z}}(1))' \quad \ldots, vec(\Omega^d_{\Delta^{\rho_z}}(L))')'$;

3: For a given ρ_z and Θ^-, solve the DT model to yield $\widehat{m}_t$. Use

$$\Delta^{\rho}\widehat{m}_t = \Pi_{DT}\Delta^{\rho}\widehat{m}_{t-1} + B_{DT}\Delta^{\rho_z}u^z_t \tag{6}$$

to compute $\Omega^m_{\Delta^{\rho_z}}(j), j = 1, \ldots L$, the model implied covariance and autocovariance matrices of the quasi-differenced variables. Let

$$\Omega^m_{\Delta^{\rho_z}}(j) = \widehat{\Omega}^m_{\Delta^{\rho_z}}(j) - \widehat{\Omega}^m_{\Delta^{\rho_z}}(0)$$

Define $\widehat{\omega}^m_{\Delta^{\rho_z}} = (vec(\Omega^m_{\Delta^{\rho_z}}(1))' \quad \ldots, vec(\Omega^m_{\Delta^{\rho_z}}(L))')'$;

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

The QD estimator is based on the difference between the model and the sample autocovariances, normalized by the variance, $\widehat{\Omega}^m_{\Delta^{\rho_z}}(0)$. As shown in the Appendix, this is necessary to obtain an asymptotic distribution that is well behaved. Hereafter, the non-normalized estimator where $\Omega_{\Delta^{\rho_j}}(j) = \widehat{\Omega}_{\Delta^{\rho_j}}(j)$ will be denoted QD^0. 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 when evaluated at the true ρ_z, whether or not the true ρ_z is inside or on the unit circle. 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 $|\rho_z| \leq 1$.

At this point it is useful to relate our approach with other methods considered in the literature. Fukac and Pagan (2006) propose using Beveridge-Nelson decomposition to estimate and remove permanent component in the data series. Apart from the fact that the permanent component in the Beveridge-Nelson decomposition may be different from actual trend and is subject to stringent assumptions, the clear advantage of our approach is that it is a one-step procedure that can handle multiple I(1) shocks.

In a study closely related to ours, Cogley (2001) investigates how an inappropriate choice of trend can lead to strong biases in the parameter estimates. He considers several alternative estimation strategies and finds that using cointegration relationships in unconditional Euler equations works quite well, as the moments used in GMM estimation remain stationary irrespective of whether the data are trend or difference stationary. Our method is similar to Cogley's (2001) in that neither requires the researcher to take a stand on the properties of the trend dynamics before estimation, but there are important differences. 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 such as adjustment costs impossible because adjustment costs are zero by construction in the 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.

3.2 The ΔDT Estimator

If the 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 the following estimation procedure:

1: First difference the observed data to obtain $\Delta d_t^c = (c_t^c, k_t^c, l_t^c)$ where

$$c_t^c = \Delta c_t - \bar{g}, \quad k_t^c = \Delta k_t - \bar{g}, \quad l_t^c = \Delta l_t. \tag{7}$$

2: Compute $\widehat{\Omega}_\Delta^d(j)$ the autocovariance matrix of Δd_t^c at lag j. Define

$$\widehat{\omega}_\Delta^d = (vech(\Omega_\Delta^d(0))' \;\; vec(\Omega_\Delta^d(1))' \;\; \ldots \; vec(\Omega_\Delta^d(L))')'.$$

3: For a given Θ, solve the DT model. Use the representation

$$\Delta\widehat{m}_t = \Pi_{DT}\Delta\widehat{m}_{t-1} + B_{DT}\Delta u_t^z \tag{8}$$

to compute $\Omega_\Delta^m(j)$, the model implied autocovariance matrices at lag j of the first differenced variables. Define

$$\widehat{\omega}_\Delta^m = (vech(\Omega_\Delta^m(0))' \;\; vec(\Omega_\Delta^m(1))' \;\; \ldots \; vec(\Omega_\Delta^m(L))')'.$$

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

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 correctly specified only when $\rho_z = 1$. Note that the QD estimator and ΔDT estimator are equivalent when $\rho_z = 1$.

3.3 The Hybrid Estimator

One drawback of the ΔDT estimator is that when ρ_z is less than unity, over-differencing induces a non-invertible moving-average component. The estimates obtained by matching a small number of lagged autocovariances may be inefficient. The QD estimator does not have this problem but it is based on the second moments of the quasi-differenced variables which themselves depend on ρ_z. As the Jacobian matrix is a function of ρ_z, the QD objective function can be highly non-linear in ρ_z and the QD estimates can be computationally more difficult to obtain. The above considerations suggest a hybrid estimator:

1: Transform the observed data to obtain $\Delta^{\rho_z} d^c_t$ (as in QD) and Δd^c_t (as in ΔDT).

2: Compute $\widehat{\Omega}^d_{QD,\Delta}(j) = cov(\Delta^{\rho_z} d^c_t, \Delta d^c_{t-j})$, the covariance between $\Delta^{\rho_z} d^c_t$ and Δd^c_{t-j}. Define

$$\widehat{\omega}^d_{QD,\Delta} = (vec(\Omega^d_{QD,\Delta}(0))' \;\; vec(\Omega^d_{QD,\Delta}(1))' \;\; \ldots \;\; vec(\Omega^d_{QD,\Delta}(L))')'.$$

3: For a given Θ, solve the DT model. Use the solution to compute model implied covariance and autocovariance between the quasi-differenced and the first differenced variables. Define

$$\widehat{\omega}^m_{QD,\Delta} = (vec(\Omega^m_{QD,\Delta}(0))' \;\; vec(\Omega^m_{QD,\Delta}(1))' \;\; \ldots \;\; vec(\Omega^m_{QD,\Delta}(L))')'.$$

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

We denote this estimator with HD (hybrid differencing). HD$\widehat{m}_t$ means that one uses both the quasi and the first differenced data to construct moments.

3.4 The HP-HP Estimator

The final robust method is based on commonly used symmetric filters such as the HP and bandpass filters, see Baxter and King (1999). A desirable feature of these filters is that they can remove deterministic as well as stochastic trends. As discussed in King and Rebelo (1993), the data can be rendered stationary without the user deciding a priori the specific type of non-stationarity that is to be handled.

The HP filter is heavily used in empirical analysis, and a HP detrended series is defined as

$$\begin{aligned} d_t^c &= HP(L)d_t = \frac{\lambda(1-L)^2(1-L^{-1})^2}{1+\lambda(1-L)^2(1-L^{-1})^2}d_t \\ &= HP^+(L)\Delta d_t = \frac{\lambda(1-L)(1-L^{-1})^2}{1+\lambda(1-L)^2(1-L^{-1})^2}\Delta d_t. \end{aligned}$$

As seen earlier, the common practice of estimating either the DT or the ST model using HP filtered data can lead to substantial bias in the parameter estimates. The reason is that the HP filter changes the autocovariance structure of the data. It follows that if we were to filter the data, we would also need to simultaneously HP filter the model variables. The procedure can be summarized as follows:

1: Let $\breve{d}_t^c$ be the HP filtered data. Compute $\Omega_{HP}^d(j)$, the autocovariance matrix at lag j of $\breve{d}_t^c$. Define

$$\widehat{\omega}_{HP}^d = (vech(\Omega_{HP}^d(0))' \;\; vec(\Omega_{HP}^d(1))' \;\; \ldots vec(\Omega_{HP}^d(L))')'.$$

2: For a given guess of Θ, solve the DT model for $\widehat{m}_t$. Compute $\Omega^m(j), j = 1, \ldots M$, the autocovariances of $\widehat{m}_t$. Apply the fourier transform to obtain the spectrum for $\widehat{m}_t$ at frequencies $2\pi j/T$, $j = 0, \ldots T-1$. Multiply the spectrum by the gain of the HP filter. Inverse Fourier transform to obtain $\Omega_{HP}(j)$, the autocovariances of the HP(L)$\widehat{m}_t$. Define

$$\widehat{\omega}_{HP}^m = (vech(\Omega_{HP}^m(0))' \;\; vec(\Omega_{HP}^m(1))' \;\; vec(\Omega_{HP}^m(L))')'.$$

4: Find the structural parameters $\widehat{\Theta} = \arg\min_{\Theta} \left\| \widehat{\omega}_{HP}^d - \omega_{HP}^m(\Theta) \right\|$.[3]

In practice, we have found that using $HP^+(L)$ and the autocovariances for Δd_t to give more stable results when ρ_z is close to one. By construction of the HP filter, both $\breve{d}_t$ and $\breve{m}_t$ are stationary for all $|\rho_z| \leq 1$.

The four robust methods can be summarized as follows:

True	Solved	Filters Used	Estimator
ST, DT	DT	Δ^{ρ_z}	QD
ST, DT	DT	Δ	ΔDT
ST, DT	DT	Δ^{ρ_z}, Δ	Hybrid (HD)
ST, DT	DT	HP	HP-HP

[3] We experimented with a simulation procedure. For each Θ, we simulated the model to generate $j = 1, ..., N$ samples of size T. For each j we computed moments. Then we averaged moments over j and used this average for ω_{HP}^m. This procedure is much slower and the results are very similar to the procedure we present in the text.

The four robust methods do not require the researcher to take a stand on whether $\rho_z < 1$ or $\rho_z = 1$ before estimation. The ST and DT are nested within QD, ΔDT and HP-HP framework.

4 Simulations

4.1 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). Since for now we have only one shock in the model, we set the standard deviation of e_t^z to $\sigma = 1$ without loss of generality. We perform 2,000 replications for each choice of parameter values. For each replication, we create series with T=200 observations which is a typical sample size in applied macroeconometric analysis.

In all simulations and for all estimators, we set starting values in optimization routines equal to the true parameter values. As is common in covariance structure estimation, we use an identity weighting matrix in our method of moments estimator. The model is solved using the Anderson and Moore (1985) algorithm. A rational expectations solution is said to be stable if the number of unstable eigenvalues of the system equals the number of forward looking variables. Stability in this context refers to the internal dynamics of the system to return to the steady state which can grow over time. This is distinct from covariance stationarity of the time series data, which in the base case model obtains when $\rho_z < 1$. It is possible for ρ_z to be mildly explosive and yet the system has a stable, unique rational expectations equilibrium. We admit such mildly explosive estimates as solutions for otherwise $\widehat{\rho}_z$ will be truncated to the right at one, making the distribution of $\widehat{\rho}_z$ highly skewed. We do, however, restrict $\widehat{\rho}_z$ to include only values consistent with a unique rational expectations equilibrium.

4.2 Results for the Baseline Model

We report simulation results for the baseline one-shock model in Table 2. The last five columns, labeled, ΔDT, QD^0, QD, HD, HP/HP corresponds to the new estimators considered. The first four columns are results for the non-robust estimators and are labeld as (XX,YY), XX stands for the filter used to compute the autocovariances of the data, while YY stands for filtered model variables from which analytical covariances are computed. Thus, (LT, $\widehat{m}_t$) means that the sample autocovariances are computed for the linearly detrended data, and model autocovariances are computed for $\widehat{m}_t$ with $|\rho_z| \leq 1$. The DGP is given in the first column.

Our simulations suggest that the estimates reported in columns (5)-(8), which correspond to the QD, unconstrained ΔDT, the hybrid (HD), and the HP-HP estimators respectively are generally centered at the true values. These point estimates are quite stable as ρ_z approaches one, though the HP-HP estimates are noticeably more variable. 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.

Consider first the (LT,$\widehat{m}_t$) combination reported in column (1), Table 2. For small to moderate values of ρ_z, the parameter estimates are centered at true values. However, there is a significant upward bias in $\widehat{\alpha}$ that increases with ρ_z to the point that at $\rho_z = 1$, the mean of $\widehat{\alpha}$ is close to one. Similar bias can be observed for $\widehat{\sigma}$. The estimates of ρ_z tend to be relatively close to true the 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.

The case of $\rho_z = 1$ is of empirical relevance because technology shocks tend to be highly persistent. As seen from Table 1, linear detrending is commonly used in estimation of DSGE models. But results in the last row of the (LT,$\widehat{m}_t$) suggest linear detrending data with stochastic trends can lead to extremely strong biases in the estimates of the structural parameters. In a univariate setting, Nelson and Kang (1981) showed that projecting a series with a unit root on time trend can lead to spurious cycles. Here, the problem can be due to the fact that the sample moments are highly collinear when there is a unit root. This lack of variation in the moments create a problem for identification of the parameters.

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 $\widehat{\sigma}$.[4] Taken at face value, these estimates suggest a significant role for capital as a mechanism for propagating shocks in the model. Results of Cogley and Nason (1995), King and Rebelo (1993) and Harvey and Jaeger (1993) indicate that the HP filter changes not only the persistence of the series but also the relative volatility and serial correlation of the series. This translates into biased estimates of all parameters because the estimator is forced to match the properties of the filtered data that are inconsistent with the model variables.

[4]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.

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 seem to be poorly determined. However, the very large standard deviations simply reflect that the estimates converge to two "corner" solutions. One solution is associated with $\widehat{\alpha} = 0.01$ and $\widehat{\sigma} \approx 8$ and the other solution is $\widehat{\alpha} = 0.99$ and $\widehat{\sigma} \approx 0.05$. Consider the latter solution which occurs in three quarters of the simulations. Note that in the ST model, shocks to $\widetilde{m}_t$ are transitory. Thus, the endogenous variables such as consumption adjust quickly to the permanent technology shock. But the HP filtered data are serial correlated. Thus, the estimator is forced to produce parameter values that can generate a stronger serial correlation in cyclical behavior of the model variables.

To get a sense of the consequence of over-differencing, consider the combination (FD,$\Delta^1\widehat{m}_t$), reported in column 4, Table 2. While the estimates are fairly precise when ρ_z is indeed equal to one, as ρ_z departs from one, Problem (MTS) manifests in an upward bias in $\widehat{\sigma}$. Even though the estimates based on (FD,$\Delta^1 m_t$) exhibit sizable biases when ρ_z moves away from one, (FD,$\Delta^1 m_t$) clearly dominates (HP,$\widetilde{m}_t$). This pattern is typical in our simulations.

4.3 Spurious Propagation Mechanisms

Clearly, the large estimates of α will alert the researcher that the model is likely misspecified and he or she must make adjustments to the model. One possible and popular modification is to introduce serial correlation in the growth rate of shocks to technology. Specifically, one might estimate the following process for u_t : $u_t = (\rho_z + \kappa)u_{t-1} - \kappa\rho_z u_{t-2} + e_t^z$. This specification generates serial correlation κ in growth rate of technology when $\rho_z \approx 1$. Our baseline model corresponds to $\kappa = 0$. When we simulate data with $\kappa = 0$ and allow this more general specification for u_t to be estimated, (QD,$\Delta^{\rho}\widehat{m}_t$), (HD,HD$\widehat{m}_t$), (FD,$\Delta\widehat{m}_t$), and (HP,HP$\widehat{m}_t$) correctly find that $\kappa = 0$ and estimates of other structural parameters are unaffected. The non-robust methods yield estimates of α around 0.4-0.5, which seem more plausible than when κ was assumed zero. However, these estimates are achieved by having $\widehat{\kappa}$ large and statistically significant even when the true κ is zero. We report these results in Panel A, Table 3. The non-robust methods (LT,$\widehat{m}_t$), (HP,$\widehat{m}_t$) and (HP,$\widetilde{m}_t$) tend to generate negative and significant estimates of κ. For (FD,$\Delta^1\widehat{m}_t$), a strong negative κ is estimated when ρ_z is much less than one. This negative correlation is necessary to dampen the responses of endogenous variables and make them less persistent.

Internal habit in consumption is another popular way to introduce greater persistence in business cycle models. Specifically, consider an alternative utility function:

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

where ϕ measures the degree of habit in consumption. We set $\phi = 0$ and estimate $(\alpha, \phi, \rho, \sigma)$ to investigate how the treatment of the trends affects estimates of internal propagation mechanisms. We report results in Panel B, Table 3.

The robust estimators perform well for all values of ρ_z. The bias in the estimates is generally negligible. This is not the case for the other estimators. The combination (LT,$\widehat{m}_t$) has a downward bias in the estimates of ϕ, but the large standard deviation of the estimates suggest that the distribution of the estimates is fairly flat. In contrast, (HP,$\widehat{m}_t$) and (HP,$\widetilde{m}_t$) have a strong upward bias in the estimates of ϕ. As we discussed above, HP filtering retain substantial serial correlation in the data and a strong habit formation is necessary to capture this persistence. Finally, (FD,$\Delta^1\widehat{m}_t$) produces a negative bias in $\widehat{\phi}$ when ρ_z departs from one. This "negative" habit formation serves to reduce persistence and increases volatility which is necessary to match the properties of the data when ρ_z is is one.

In both cases considered, the fit of the misspecified models improves relative to the correctly specified model. However, these modifications should not have been undertaken as they do not exist in the data generating process. These examples indicate how the treatment of trends can motivate the researcher to augment correctly specified models with spurious propagation mechanisms to match the moments of the data.[5] The estimates can yield misleading inference about the relative importance of habit and persistence of technology shocks.

4.4 Statistical Properties

The above simulations indicate that the robust estimates are close to the true value. Recall that ω^d are the sample moments of the filtered variables while $\omega^m(\Theta)$ are the model implied moments based on the same filter. Define

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

Thus, $\widehat{\Theta} = \min_\Theta \bar{g}(\Theta)' g(\Theta)$ is a non-linear GMM estimator with an identity weighting matrix. Let G be the matrix of derivatives of $\bar{g}$ with respect to Θ. Standard asymptotic theory suggests that

$$\sqrt{T}(\widehat{\Theta} - \Theta_0) \xrightarrow{d} A \cdot N(0, S)$$

where $A = (G_0' G_0)^{-1} G_0'$ and $\sqrt{T}\bar{g}(\Theta_0) \xrightarrow{d} N(0, S)$. It remains to address whether our robust estimators have these properties.

To shed some light on the large sample properties of the estimators, Figure 1 presents the standardized distribution of $\widehat{\alpha}$, denoted $t_{\widehat{\alpha}}$ when $\rho_z = 0.95$, with the normal distribution super-imposed.

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

Figure 2 presents the results for $\rho_z = 1.0$. These t-statistics are computed for $T = 1000$ using Newey-West standard errors. Notably, the linear detrending yields strongly biased estimates. The sampling distributions for the standardized robust estimators appear symmetric. The approximate normality of the finite sample distributions, especially that of $\widehat{\rho}_z$ when ρ_z is so close to one is totally unexpected, given that the literature on integrated regressors prepared us to expect super consistent estimators with Dickey-Fuller type distributions that are skewed.

An important feature of our estimators is that the sample moments are stationary when evaluated at the true parameter vector. This permits application of central limit theory. When $\rho_z < 1$, the estimators are simply $\sqrt{T}$ consistent and asymptotically normal. Whether the estimator is normal or is mixed normal when $\rho_z = 1$ depends whether G_0 is random or not. For the FD and the HP-HP, G_0 has non-random limits. The estimators are thus $\sqrt{T}$ consistent and asymptotically normal. For the hybrid estimator, G_0 is random when $\rho_z = 1$. Thus, the matrix A in the asymptotic distribution is non-random only if the data are stationary, so that the hybrid estimator is mixed normal when $\rho^z = 1$. However, the t statistics remain approximately normal.

Turning to the QD and QD^0, note first that when $\rho^z = 1$, the quasi-transformed data remain non-stationary except when the filter $1 - \rho^z L$ is evaluated at the true ρ^z. This has implications for the behavior of $\bar{g}$ in the neighborhood of the true ρ^z of 1. In particular, recall that the QD^0 uses as $\bar{g}$ the difference in variance and autocovariances between the data and the model. The quadratic term in the expansion of $\bar{g}(\Theta)$ is $O_p(T)$ instead of $O_p(1)$. Because of this, we have been unable to derive the asymptotic distribution of the QD^0, even though it may be well defined. We can, however, obtain the asymptotic distribution of the QD, which uses as $\bar{g}$ the autocovariances at lag j normalized by the variance. In a nutshell, the ill-behaved terms in each autocovariance at lag j and that in the variance cancel out, resulting in the rather unexpected property that $\widehat{\Theta}$ is asymptotically normal. A sketch of the argument for the QD is given in the Appendix for the model with inelastic supply, because the closed-form solution for that model is known. The properties of the robust estimators are further investigated in Gorodnichenko and Ng (2007) in the context simpler linear regression models.

From a practical perspective, the primary advantage of the robust estimators is that when properly studentized, the estimators are normally distributed whether or not ρ^z is less than one, which greatly facilitates inference. It should also be noted that the estimators except the QD are linear in ρ_z; non-linear estimation for the FD, Hybrid, and HP-HP is necessary only because of cross-equation constraints. The QD is genuinely a non-linear estimator, and it estimates ρ^z through variation in the quasi-transformed sample moments. With the other estimators, the moments of the filtered data are unaffected by ρ^z. Which

5 The General Formulation

Extension of the robust estimators to more general cases is straightforward. Let m_t be a vector of q (predetermined, non-predetermined, plus exogenous) model variables and let $\widehat{m}_t$ be the vector of zero-mean variables that are deviations of m_t from the steady state values. Let d_t be a vector of r observed variables. The general solution in state space representation is

$$d_t = \delta_0 + \delta_1 t + B\,\widehat{m}_t \tag{9}$$

The measurement equation (9) links d_t to the q model variables $\widehat{m}_t$ through the matrix B $(r \times q)$.[6] The parameter vectors δ_0 and δ_1 are restricted constants that can be estimated along with the other parameters. Or, we can linearly detrend the data prior to estimation as in Ireland (2004a). Then $\widehat{d}_t = d_t - \widehat{\delta}_0 - \widehat{\delta}_1 t$ is the detrended data. If the model is correctly specified for the data, both methods of detrending are asymptotically equivalent. From

$$\widehat{d}_t = d_t - \widehat{\delta}_0 - \widehat{\delta}_1 t = B\,\widehat{m}_t \tag{10}$$

we can define $d_t^c = \Delta\widehat{d}_t - \delta_1 = B\,\Delta\widehat{m}_t$ if the ΔDT estimator is to be used. Implementation of the HP-HP filter can be obtained upon replacing the first difference filter by the HP filter. The important point is that whatever filter we apply to $\widehat{d}_t$ to arrive at a d_t^c that is stationary, the same filter should be applied to $\widehat{m}_t$. Estimation then proceeds by matching the moments of the filtered model variables and the filtered data.

Extension of the estimators to handle multiple shocks is straightforward. Suppose there are J shock processes u_{jt} with

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

where some of the ρ_j may be on the unit circle. Define

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

Now the quasi-differencing operator is the product of the J polynomials in lag operator that describes the dynamics of the J shocks. Once the DT model is solved, we can compute moments for the quasi-differences of $\widehat{m}_t$. Whether none, one, or both shocks are permanent, the autocovariances of the transformed variables are well defined. It is also possible to perform partial quasi-differencing. For example, if one knows that shocks to tastes dissipate quickly while technology shocks z_t are highly persistent, the researcher can use only $(1 - \rho_z L)$ in the Δ^ρ operator.

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

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

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

In this model, the parameters are $(\alpha, \rho_z, \sigma_z, \rho_q, \sigma_q)$. We fix $\sigma_z = 1$ and let σ_q take values (0.5,1,1.5). We set $\rho_q = 0.8$ and vary ρ_z from 0.95 to 1 so that in our exercise technology shocks are generally more persistent than shocks to hours. To preserve space, we report only selected results for $\widehat{\alpha}$ in Table 4. Additional results are available upon request.

The robust estimators perform well for all values of ρ_z. For the non-robust methods, The combination (FD,$\Delta^1 m_t$) performs well when technology shocks have a unit root but not when ρ_z is far from one. The (LT,$\widehat{m}_t$) combination produces imprecise estimates throughout and the biases are more pronounced when shocks to hours become more persistent. The combination (HP, $\widehat{m}_t$) has larger biases than (HP, $\widetilde{m}_t$), which seems to be sensitive to σ_q. In some cases, the non-robust estimates suggest $\widehat{\sigma}_q > \widehat{\sigma}_z$, 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.

To compare the properties of the estimators, Figures 3 to 5 show $T\times$MSE of the five estimators for the baseline model for T ranging from 100 to 2000. Several features are of note. The estimators are evidently consistent, since T times the MSE is fairly constant. For all three parameters considered, the QD is the most efficient of the five estimators. The estimates produced by the HP and the ΔDT have much larger MSE, especially for $\widehat{\rho}$. Although the QD^0 performs quite well, we cannot establish that its distribution is asymptotic normal, as in the case of the QD. Figure 6 ($\rho^z = .95$) and 7 ($\rho^z = 1.0$) show $T \times MSE$ for (i) the model that estimates the habit parameter when it is in fact zero, (ii) the model that allows for correlated shocks when they are in fact uncorrelated, and (iii) the two shock model. To conserve space, we only report results for $\widehat{\alpha}$. The HP again has the largest while the QD has smallest errors. The robustness of the QD across models and parameters make it appealing in empirical work.

6 Concluding Remarks

This paper has several substantive findings. First, we show that Problems (DD) and (MTS) can lead to distorted estimates. Researchers can be misled to non-existent propagating mechanisms to be statistically significant. Second, we consider robust approaches that essentially let the data decide whether the model should specify deterministic or stochastic trends. This is done by transforming the model variables and the data using the same filter, and the filter must produce stationary series when evaluated at the parameter vector. We consider four such transformations, though other methods such as band-pass filter can also be used. As the filtering takes place within the state-space framework, likelihood based estimation can also be used in place of the covariance estimator considered here. We leave such an analysis for future research.

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.

Anderson, G. and Moore, G. 1985, A linear algebraic procedure for solving linear perfect foresight models, *Economic Letters* **17(3)**, 247–252.

Baxter, M. and King, R. 1999, Measuring Business Cycles: Approximate Bandpass Filters for Economic Time Series, *Review of Economics and Statistics* **81:4**, 575–593.

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. and Eichenbaum, M. 1996, Small-sample Properties of GMM-based Wald Tests, *Jornal of Business and Economic Statistics* **14**, 294–308.

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.

Fernandez-Villaverde, J. and Rubio-Ramirez, J. F. 2006, Estimating Macroeconomic Models: A Likelihood Approach, University of Pennsylvania, mimeo.

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.

Fukac, M. and Pagan, A. 2006, Limited Information estimation and Evaluation of DSGE Models, NCER Working Paper No. 6.

Gorodnichenko, Y. and Ng, S. 2007, Estimators for Persistent and Possibly Non-Stationary Data with Classical Properties, mimeo.

Gregory, A. and Smith, G. 1996, Measuring Business Cycles with Business-cycle Models, *Jornal of Economic Dynamics and Control* **20**, 1007–1025.

Harvey, A. and Jaeger, A. 1993, Detrending, Stylized Factors, and the Business Cycle, *Jornal of Applied Econometrics* **8**, 231–247.

Ireland, P. 1997, A Small Structural Quarterly Model for Monetary Policy Evaluaton, *Carnegie-Rochester Conference Series on Public Policy* **47**, 83–108.

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, mimeo.

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.

7 Appendix

To study the analytical properties of the QD, we consider the one sector model with inelastic labor supply, since its closed form solution is known.

$$\begin{aligned}
k_t &= v_{kk}k_{t-1} + v_{kz}z_t \\
r_t &= v_{rk}k_{t-1} + v_{rz}z_t \\
c_t &= v_{ck}k_{t-1} + v_{cz}z_t \\
y_t &= z_t + \alpha k_{t-1} \\
z_t &= \rho z_{t-1} + e_t
\end{aligned}$$

The endogenous variables in this model are ARMA(2,1) processes. To see this, focus on y. From

$$\begin{aligned}
y_t &= z_t + \alpha k_{t-1} \\
y_t &= z_t + \alpha \frac{v_{kz}z_{t-1}}{1 - v_{kk}L} \\
y_t &= v_{kk}y_{t-1} + z_t - v_{kk}z_{t-1} + \alpha v_{kz}z_{t-1} \\
&= v_{kk}y_{t-1} + z_t(1 - aL) \\
&= v_{kk}y_{t-1} + \frac{(1-\theta_y L)}{1-\rho L}e_t = \frac{1-\theta_y L}{(1-v_{kk}L)(1-\rho L)}e_t \\
y_t(1 - v_{kk}L)(1-\rho L) &= (1-\theta_y L)e_t.
\end{aligned}$$

Thus y_t is an ARMA(2,1) where $\theta_y = v_{kk} - \alpha v_{kz}$. Note that v_{kk} does not depend on ρ, but v_{kz} depends on ρ.

Similarly, for consumption,

$$\begin{aligned}
c_t &= v_{ck}k_{t-1} + v_{cz}z_t \\
&= v_{kk}c_{t-1} + v_{cz}z_t - v_{cz}v_{kk}z_{t-1} + v_{ck}v_{kz}z_{t-1} = \frac{1-\theta_c L}{(1-v_{kk}L)(1-\rho L)}v_{cz}e_t.
\end{aligned}$$

Let $\theta_c = v_{kk} - \frac{v_{ck}}{v_{cz}}v_{kz}$. Then c_t is also an ARMA(2,1):

$$c_t(1-v_{kk}L)(1-\rho L) = (1-\theta_c L)v_{cz}e_t.$$

The moving average representation for an ARMA(2,1) (assuming $\rho > v_{kk}$)

$$\begin{aligned}
\psi_0 &= 1 \\
\psi_1 &= \theta + \rho \\
\psi_j &= \psi_{j-1}\rho^{j-1} + \psi_{j-2}v_{kk}, \quad j \geq 2.
\end{aligned}$$

Generically, write the DGP as

$$(1-\rho^0 L)(1 - v^0_{kk}L)y_t = e_t + b(\rho^0, v^0_{kk})e_{t-1}$$

where $b(\rho, v_{kk})$ is continuous in ρ, v_{kk}. Let $\widetilde{y}^0_t = (1-\rho^0 L)y_t$ be the data quasi-differenced at the true ρ. Then

$$\widetilde{y}^0_t = v^0_{kk}\widetilde{y}^0_{t-1} + e_t + b(\rho^0, v^0_{kk})e_{t-1}$$

is an ARMA(1,1) with autocovariances defined by

$$\begin{aligned}\widetilde{\gamma}_0(\rho^0) &= \sigma^2 \frac{1+b(\rho^0,v_{kk}^0)^2+2v_{kk}^0 b(\rho^0,v_{kk}^0)}{1-v_{kk}^2} \\ \widetilde{\gamma}_1(\rho^0) &= \sigma^2 \frac{(1+v_{kk}^0 b(\rho^0,v_{kk}^0))(v_{kk}^0+b(\rho^0,v_{kk}^0))}{1-v_{kk}^2} \\ \widetilde{\gamma}_j(\rho^0) &= v_{kk}^{0,j-1}\widetilde{\gamma}_1(\rho^0), \quad j\geq 2\end{aligned}$$

The derivatives of $\widetilde{\gamma}_j(\rho)$ with respect to ρ, σ^2, and v_{kk} exist and are well defined. To focus on the issue, we fix σ^2 and v_{kk} in the derivations to follow.

Let $\widetilde{\gamma}_{Tj}(\rho)$ be the sample autocovariance of the y_t quasi-transformed at an arbitrary ρ. By direct calculations,

$$\widetilde{\gamma}_{Tj}(\rho) = \frac{1}{T}\sum_{t=1}^{T} \widetilde{y}_t^0\widetilde{y}_{t-j}^0 - (\rho-\rho^0)y_{t-1}\widetilde{y}_{t-j}^0 - (\rho-\rho^0)y_{t-j-1}\widetilde{y}_t^0 + (\rho-\rho^0)^2 y_{t-1}y_{t-j-1}.$$

This implies

$$\widetilde{\gamma}'_{Tj}(\rho) = \frac{\partial\widetilde{\gamma}_{Tj}(\rho)}{\partial\rho} = -\frac{1}{T}\sum_{t=1}^{T}\left(y_{t-1}\widetilde{y}_{t-j}^0 + y_{t-j-1}\widetilde{y}_t^0\right) + 2\frac{1}{T}(\rho-\rho^0)\sum_{t=1}^{T} y_{t-1}y_{t-j-1} \tag{11}$$

$$\widetilde{\gamma}''_{Tj}(\rho) = \frac{\partial^2\widetilde{\gamma}_{Tj}(\rho)}{\partial\rho^2} = \frac{2}{T}\sum_{t=1}^{T} y_{t-1}y_{t-j-1} \tag{12}$$

Furthermore, $\frac{\partial\widetilde{\gamma}_{Tj}(\rho)}{\partial v_{kk}} = 0$. Let

$$g_j(\rho) = \widetilde{\gamma}_{Tj}(\rho) - \widetilde{\gamma}_j(\rho)$$

be the difference between the sample and model autococovariance at lag j. The corresponding gradient and Hessian are

$$\begin{aligned}G_{Tj}(\rho) &= \frac{\partial\widetilde{\gamma}_{Tj}(\rho)}{\partial\rho} - \frac{\partial\widetilde{\gamma}_j(\rho)}{\partial\rho} \\ H_{Tj}(\rho) &= \frac{\partial^2\widetilde{\gamma}_{Tj}(\rho)}{\partial\rho^2} - \frac{\partial^2\widetilde{\gamma}_j(\rho)}{\partial\rho^2},\end{aligned}$$

The quadratic expansion of $\bar{g}(\rho)$ around $\bar{g}(\rho^0)$ is

$$\bar{g}(\rho) = \bar{g}(\rho^0) + G'_T(\rho^0)(\rho-\rho^0) + \frac{1}{2}H_T(\bar{\rho})(\rho-\rho^0)^2.$$

Case 1: $\rho^0 < 0$ Note first that $\widetilde{g}_{Tj}(\rho) = O_p(1)$ and $\widetilde{g}_j(\rho) = O(1)$ for all $\rho < 1$. From (11) $G_T(\rho^0) = O_p(1)$. From (12) $H_T(\rho) = O_p(1)$. The first order condition $G_T(\widehat{\rho})\bar{g}(\widehat{\rho}) = 0$ becomes

$$\begin{aligned}0 &= G'_T(\widehat{\rho})\left(\bar{g}(\rho^0) + G_T(\rho^0)(\widehat{\rho}-\rho^0) + \frac{1}{2}H_T(\widehat{\rho}-\rho^0)^2\right) \\ &= G'_T(\widehat{\rho})\bar{g}(\rho^0) + G'_T(\widehat{\rho})G_T(\rho^0)(\widehat{\rho}-\rho^0) + G'_T(\widehat{\rho})\frac{1}{2}H_T(\bar{\rho})(\widehat{\rho}-\rho^0)^2 \\ &= G'_T(\widehat{\rho})\bar{g}(\rho^0) + G'_T(\widehat{\rho})G_T(\rho^0)(\widehat{\rho}-\rho^0) + O_p(T^{-1}).\end{aligned}$$

By assumption, $\sqrt{T}\bar{g}(\rho^0) \xrightarrow{d} N(0,S)$. Thus when $\rho < 1$,

$$\begin{aligned}
\sqrt{T}(\widehat{\rho}-\rho) &= -(G_T'(\widehat{\rho})\bar{G}(\widehat{\rho}))^{-1}G_T'(\widehat{\rho})\sqrt{T}\bar{g}(\rho^0) + o_p(1) = K_T\bar{g}(\rho^0) + o_p(1) \\
&\xrightarrow{d} K(\rho_0)\cdot N(0,S) = N(0, K_0SK_0').
\end{aligned}$$

Case 2. When $\rho^0 = 1$, we can write the DGP as

$$\begin{aligned}
\Delta y_t &= v_{kk}^0\Delta y_{t-1} + e_t + b(\rho^0)e_{t-1} \\
\Delta y_t &= \frac{e_t + b(\rho^0)e_{t-1}}{1 - v_{kk}^0 L} = u_t \\
u_t &= v_{kk}^0 u_{t-1} + e_t + b(\rho^0)e_{t-1}
\end{aligned}$$

where u_t is stationary and $2\pi f_u(0) = \omega^2$. Thus, y_t has a unit root, and the functional central limit theory holds that $\frac{1}{\sqrt{T}\omega}\sum_{j=0}^t u_j \Rightarrow W(r)$, where $W(r)$ is a standard Brownian motion.

Using $y_{t-1} = y_{t-j} + u_{t-1} + u_{t-2} + \dots u_{t-j+1}$ for $j > 1$,

$$y_{t-1}^2 = y_{t-1}y_{t-j-1} + y_{t-1}(u_{t-1} + u_{t-2} + \dots u_{t-j+1})$$

Thus, with $\widetilde{\gamma}_{T0}''(\rho) = \frac{1}{T}\sum_{t=1}^T y_{t-1}^2$ which does not depend on j or ρ. Furthermore,

$$\widetilde{\gamma}_{Tj}''(\rho) \approx \widetilde{\gamma}_{T0}''(\rho) - \frac{1}{T}\sum_{t=1}^T y_{t-1}(u_{t-1} + u_{t-2} + \dots u_{t-j-1}).$$

where $\frac{1}{T}\sum_{t=1}^T y_{t-1}(u_{t-1} + u_{t-2} + \dots u_{t-j+1}) = O_p(1)$. THus, $H_{Tj}^*(\rho) = O_p(1)$. For $j = 1, \dots L$,

$$\bar{g}_{Tj}^*(\rho) = \bar{g}_{Tj}(\rho) - \bar{g}_{T0}(\rho).$$

The first order condition is $G_T^{*\prime}(\widehat{\rho})\bar{g}^*(\widehat{\rho}) = 0$ and

$$\sqrt{T}\bar{g}^*(\widehat{\rho}) = \sqrt{T}\bar{g}^*(\rho^0) + G_T^{*\prime}(\rho^0)\sqrt{T}(\widehat{\rho}-\rho^0) + \frac{1}{2}\sqrt{T}H_T^*(\bar{\rho})(\widehat{\rho}-\rho^0)^2$$

where $\bar{\rho} \in [\widehat{\rho}, \rho^0]$. The last term is $o_p(1)$ if $\sqrt{T}(\widehat{\rho}-\rho) = O_p(1)$. Thus

$$\begin{aligned}
0 &= G_T^{*\prime}(\widehat{\rho})\Big(\sqrt{T}\bar{g}(\rho^0) + G_T^*(\rho^0)\sqrt{T}(\widehat{\rho}-\rho^0) + o_p(1)\Big) \\
G_T^{*\prime}(\widehat{\rho})G_T^*(\rho^0) &= -G_T^{*\prime}(\rho^0)\sqrt{T}\bar{g}^*(\rho^0) + o_p(1).
\end{aligned}$$

Thus

$$\begin{aligned}
\sqrt{T}(\widehat{\rho}-\rho^0) &= -(G_T^{*\prime}(\widehat{\rho})G_T^*(\rho^0)^{-1}G_T^{*\prime}(\rho^0)\sqrt{T}\bar{g}^*(\rho^0) + o_p(1) \\
&= K_0^*N(0,S^*) = N(0, K_0^*S^*K_0^{*\prime}).
\end{aligned}$$

Table 1: Summary of Selected Work

Paper	z_t	Equations	Filter	Estimator
Ireland (2001)	AR	system	linear trend	MLE
Del Negro et al. (2007)	unit root	system	first difference	Bayesian
Bouakez et al. (2005)	AR	system	linear trend	MLE
Faia (2007)	AR	system	HP	calibration
Clarida et al. (2000)	AR	equation	HP and CBO	GMM
Christiano et al. (2005)	not specified	system	VAR	GMM
Dib (2003)	AR	system	linear trend	MLE
Smets and Wouters (2003)	AR	system	HP	Bayesian
Smets and Wouters (2007)	AR	system	first difference	Bayesian
Kim (2000)	AR	system	linear trend	MLE
McGrattan et al. (1997)	AR	system	linear trend and HP	MLE
Altug (1989)	AR	system	First difference	MLE (freq. domain)
Fuhrer and Rudebusch (2004)	random walk	system	HP, BP, CBO, trend breaks, quadratic trends	MLE, GMM
Fuhrer (1997)	not specified	equaton	HP, linear and quadratic trends	GMM
Kydland and Prescott (1982)	permanent and transitory. shocks	system	HP	calibration
Altig et al. (2004)	unit root	system	first difference	GMM
Ireland (2004b)	random walk	system	stationary	MLE
Christiano and Eichenbaum (1992)	random walk	system	HP	GMM
Burnside et al. (1993)	AR	system	HP	GMM
Burnside and Eichenbaum (1996)	AR	system	HP	GMM

CBO denotes actual series minus the CBO measure of potential output.
unit root means $\rho^z = 1$, and Δz_t is serially correlated.
random walk means $\rho^z = 1$ and Δz_t is serially uncorrelated.

Table 2. One-shock model: Base Case.

DGP	ρ_z	XX YY	LT $\widehat{m}_t$	HP $\widehat{m}_t$	HP $\widetilde{m}_t$	FD $\Delta^1\widehat{m}_t$	ΔDT	QD^0	QD	HD	HP
			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Estimate of α											
DT	0.50	mean	0.350	0.417	0.014	0.337	0.337	0.348	0.345	0.334	0.336
		sd	0.033	0.010	0.059	0.023	0.026	0.034	0.027	0.022	0.077
DT	0.95	mean	0.477	0.676	0.794	0.398	0.366	0.323	0.323	0.336	0.349
		sd	0.121	0.023	0.392	0.082	0.110	0.051	0.053	0.061	0.103
DT	0.99	mean	0.825	0.789	0.674	0.375	0.368	0.318	0.319	0.332	0.358
		sd	0.189	0.024	0.458	0.107	0.111	0.053	0.054	0.067	0.119
ST	1.00	mean	0.906	0.818	0.603	0.351	0.333	0.317	0.315	0.320	0.347
		sd	0.190	0.022	0.479	0.110	0.088	0.055	0.054	0.062	0.110
Estimate of ρ											
DT	0.50	mean	0.452	0.248	1.000	1.000	0.497	0.476	0.483	0.495	0.479
		sd	0.075	0.051			0.103	0.064	0.070	0.091	0.142
DT	0.95	mean	0.914	0.541	1.000	1.000	0.950	0.940	0.950	0.949	0.951
		sd	0.042	0.048			0.017	0.021	0.006	0.014	0.014
DT	0.99	mean	0.864	0.486	1.000	1.000	0.990	0.988	0.989	0.990	0.991
		sd	0.094	0.041			0.007	0.008	0.002	0.005	0.011
ST	1.00	mean	0.686	0.461	1.000	1.000	0.997	1.000	0.999	0.999	1.000
		sd	0.129	0.039			0.005	0.005	0.001	0.003	0.015
Estimate of σ											
DT	0.50	mean	1.057	1.203	11.094	1.689	1.027	1.045	1.039	1.016	1.057
		sd	0.121	0.088	0.982	0.108	0.151	0.117	0.103	0.123	0.344
DT	0.95	mean	1.131	1.950	1.593	1.326	1.151	0.970	0.990	1.024	1.076
		sd	0.278	0.164	3.093	0.279	0.439	0.156	0.125	0.187	0.296
DT	0.99	mean	4.424	2.907	2.209	1.172	1.133	0.978	0.984	1.015	1.105
		sd	2.751	0.391	3.152	0.319	0.336	0.138	0.117	0.173	0.301
ST	1.00	mean	22.634	3.287	2.565	1.084	1.022	0.980	0.976	0.985	1.082
		sd	15.592	0.472	3.136	0.310	0.226	0.143	0.111	0.149	0.858

Note: The number of simulations is 2,000. Sample size is T=200. 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, HD is hybrid differencing. Δ^1 denotes the restriction $\rho_z = 1$. $\Delta^\rho = 1 - \rho_z L$ denotes quasi differencing.

Table 3. The extended One-shock model

DGP	ρ_z	XX YY	LT $\widehat{m}_t$	HP $\widehat{m}_t$	HP $\widetilde{m}_t$	FD $\Delta^1\widehat{m}_t$	ΔDT	QD^0	QD	HD	HP
			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
Panel A: estimate of κ											
DT	0.50	mean	0.075	0.111	-0.554	-0.497	0.030	-0.002	0.017	0.024	-0.007
		sd	0.163	0.038	0.025	0.038	0.154	0.203	0.194	0.148	0.315
DT	0.95	mean	-0.179	-0.224	-0.369	-0.099	-0.001	-0.006	-0.005	-0.001	-0.017
		sd	0.162	0.039	0.107	0.035	0.050	0.067	0.063	0.058	0.159
DT	0.99	mean	0.405	-0.257	-0.425	-0.021	0.004	-0.003	-0.007	0.002	-0.015
		sd	0.700	0.038	0.075	0.030	0.039	0.046	0.045	0.045	0.153
ST	1.00	mean	-0.599	-0.256	-0.447	-0.002	0.002	-0.007	-0.008	-0.001	-0.018
		sd	0.031	0.038	0.058	0.027	0.035	0.041	0.039	0.039	0.160
Panel B: estimate of ϕ											
DT	0.50	mean	-0.042	0.037	0.463	-0.325	0.029	0.075	0.027	0.061	0.057
		sd	0.222	0.281	0.050	0.039	0.118	0.269	0.205	0.192	0.473
DT	0.95	mean	-0.264	0.177	0.676	-0.087	0.002	-0.017	0.009	0.003	-0.046
		sd	0.341	0.419	0.074	0.062	0.076	0.071	0.084	0.079	0.278
DT	0.99	mean	-0.366	0.503	0.635	-0.016	0.007	0.000	0.013	0.007	0.028
		sd	0.375	0.402	0.069	0.083	0.073	0.071	0.074	0.079	0.283
ST	1.00	mean	-0.366	0.611	0.619	0.011	0.006	0.002	0.014	0.009	0.023
		sd	0.376	0.371	0.067	0.092	0.076	0.069	0.068	0.083	0.257

Note: Other parameters fixed at $\alpha = 0.33$ and $\sigma = 1$. The number of simulations is 2,000. Sample size is T=200. 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, HD is hybrid differencing. Δ^1 denotes the restriction $\rho_z = 1$. $\Delta^\rho = 1 - \rho_z L$ denotes quasi differencing.

Table 4. The Two-shock model: estimates of α.

DGP	ρ_z	XX YY	LT $\widehat{m}_t$ (1)	HP $\widehat{m}_t$ (2)	HP $\widetilde{m}_t$ (3)	FD $\Delta^1\widehat{m}_t$ (4)	ΔDT (5)	QD^0 (6)	QD (7)	HD (8)	HP
$\sigma_q = 0.5$											
DT	0.950	mean	0.465	0.591	0.337	0.299	0.361	0.348	0.344	0.335	0.329
		sd	0.126	0.037	0.137	0.043	0.080	0.051	0.040	0.025	0.075
DT	0.990	mean	0.561	0.651	0.453	0.339	0.359	0.356	0.351	0.339	0.332
		sd	0.326	0.048	0.250	0.060	0.078	0.062	0.049	0.032	0.075
ST	1.000	mean	0.508	0.665	0.485	0.352	0.352	0.361	0.353	0.342	0.335
		sd	0.357	0.050	0.265	0.063	0.067	0.067	0.052	0.034	0.072
$\sigma_q = 1.0$											
DT	0.95	mean	0.431	0.504	0.344	0.316	0.349	0.339	0.339	0.341	0.331
		sd	0.103	0.031	0.022	0.022	0.049	0.025	0.023	0.030	0.060
DT	0.99	mean	0.565	0.528	0.359	0.335	0.353	0.344	0.345	0.344	0.336
		sd	0.233	0.036	0.022	0.024	0.048	0.029	0.026	0.031	0.056
ST	1.00	mean	0.611	0.529	0.364	0.341	0.357	0.348	0.347	0.347	0.340
		sd	0.257	0.036	0.023	0.025	0.050	0.035	0.028	0.034	0.055
$\sigma_q = 1.5$											
DT	0.95	mean	0.399	0.469	0.378	0.326	0.349	0.337	0.338	0.343	0.333
		sd	0.078	0.020	0.024	0.017	0.042	0.023	0.021	0.030	0.053
DT	0.99	mean	0.474	0.477	0.387	0.336	0.353	0.341	0.341	0.345	0.336
		sd	0.171	0.021	0.023	0.017	0.040	0.021	0.020	0.030	0.050
ST	1.00	mean	0.515	0.477	0.391	0.338	0.353	0.342	0.341	0.346	0.338
		sd	0.203	0.021	0.023	0.017	0.038	0.022	0.020	0.029	0.049

Note: The true value of $\alpha = 0.33$. In all simulations, $\rho_q = 0.8$ and $\sigma_z = 1$. The number of simulations is 2,000. Sample size is T=200. 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, HD is hybrid differencing. Δ^1 denotes the restriction $\rho_z = 1$. $\Delta^\rho = 1 - \rho_z L$ denotes quasi differencing.

Figure 1: $\rho_z = 0.95$

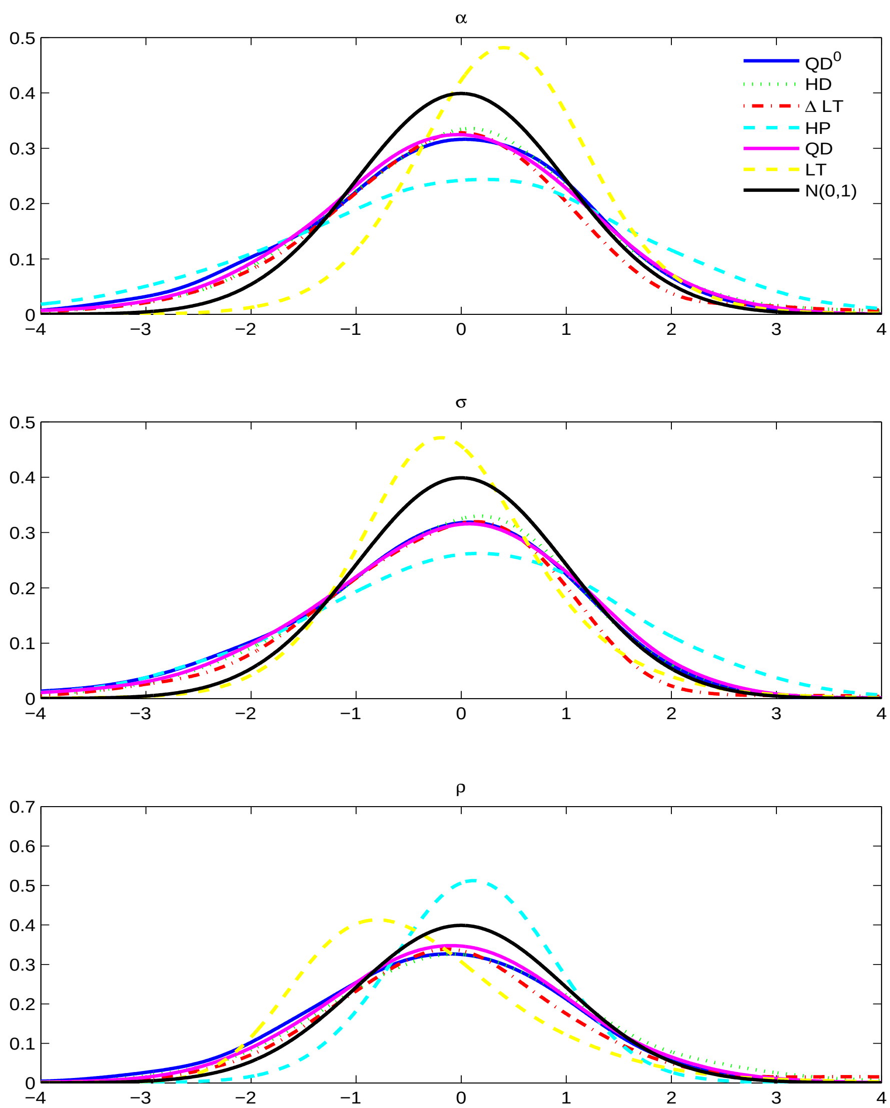

Figure 2: $\rho_z = 1.00$

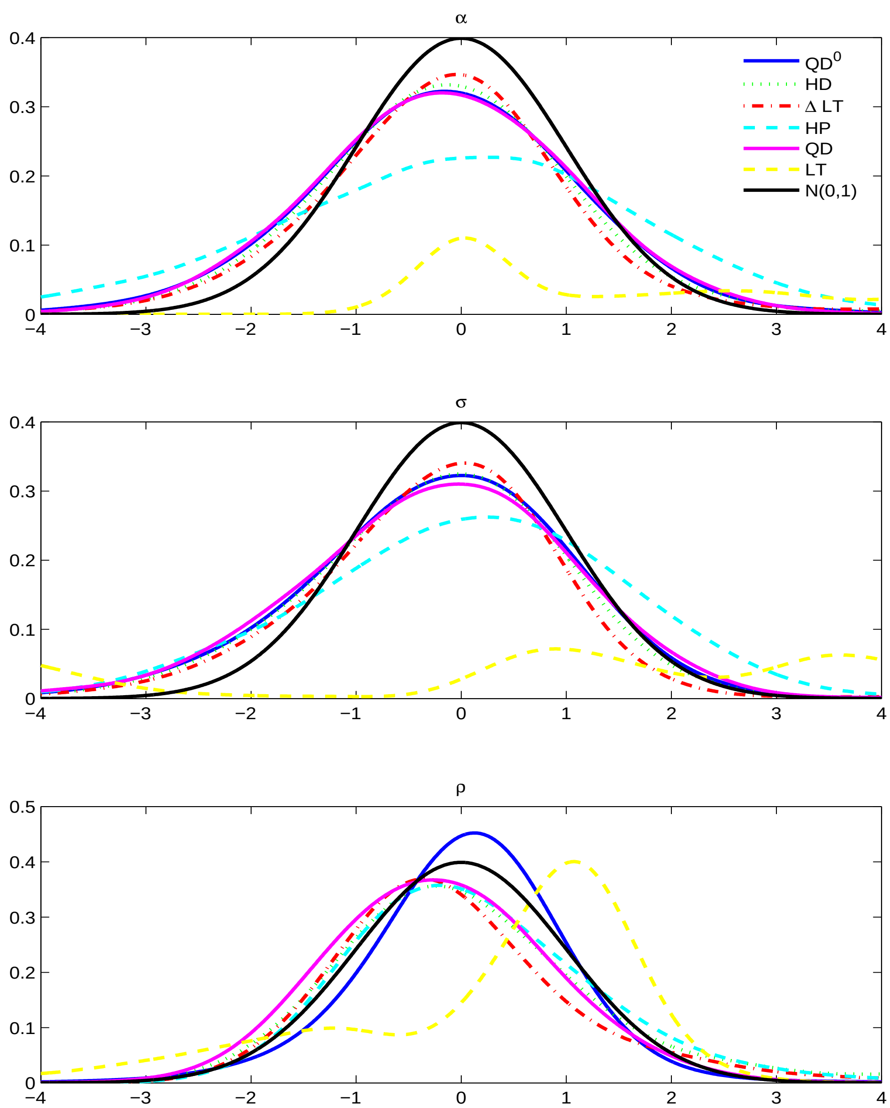

Figure 3: Base Case: $T\times$ MSE, $\rho = 0.5$

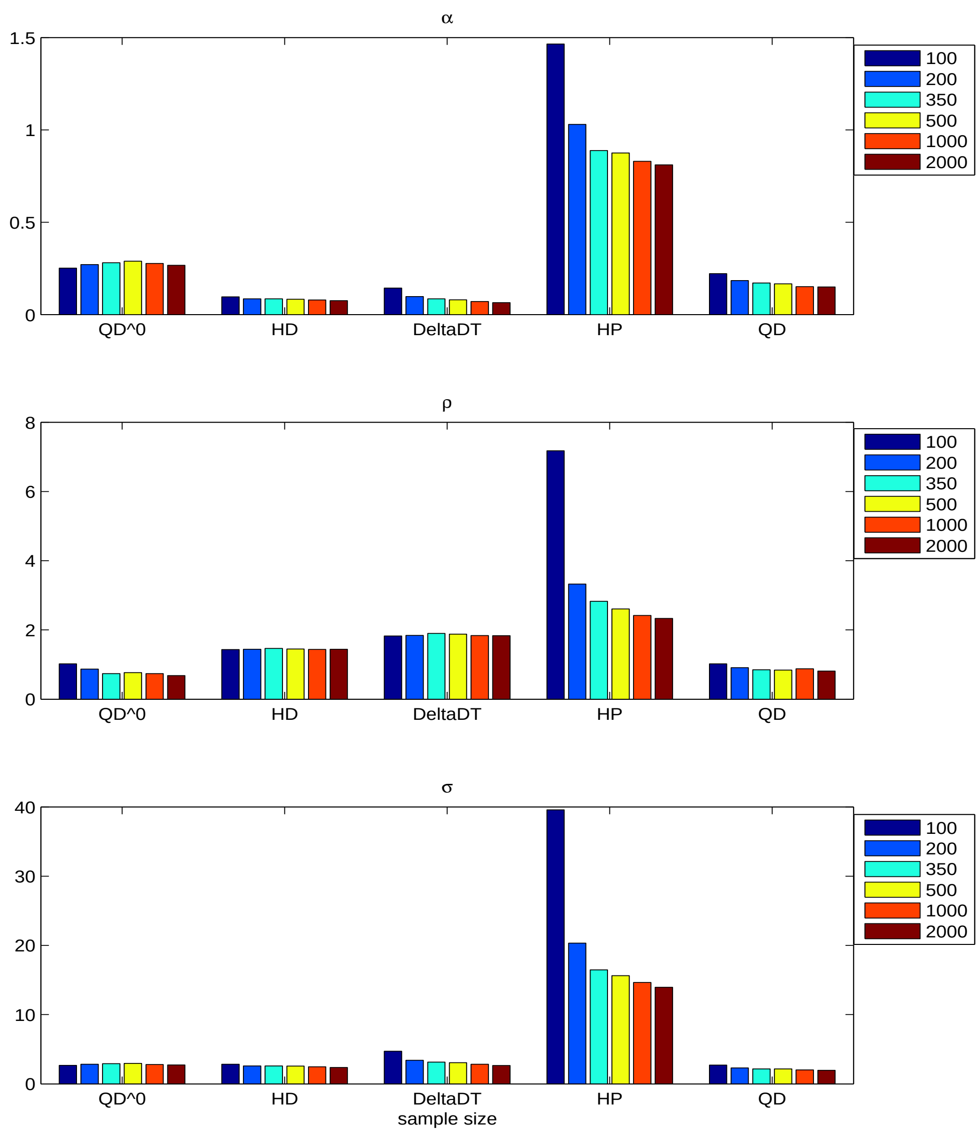

Figure 4: Base Case: $T\times$ MSE, $\rho = 0.95$

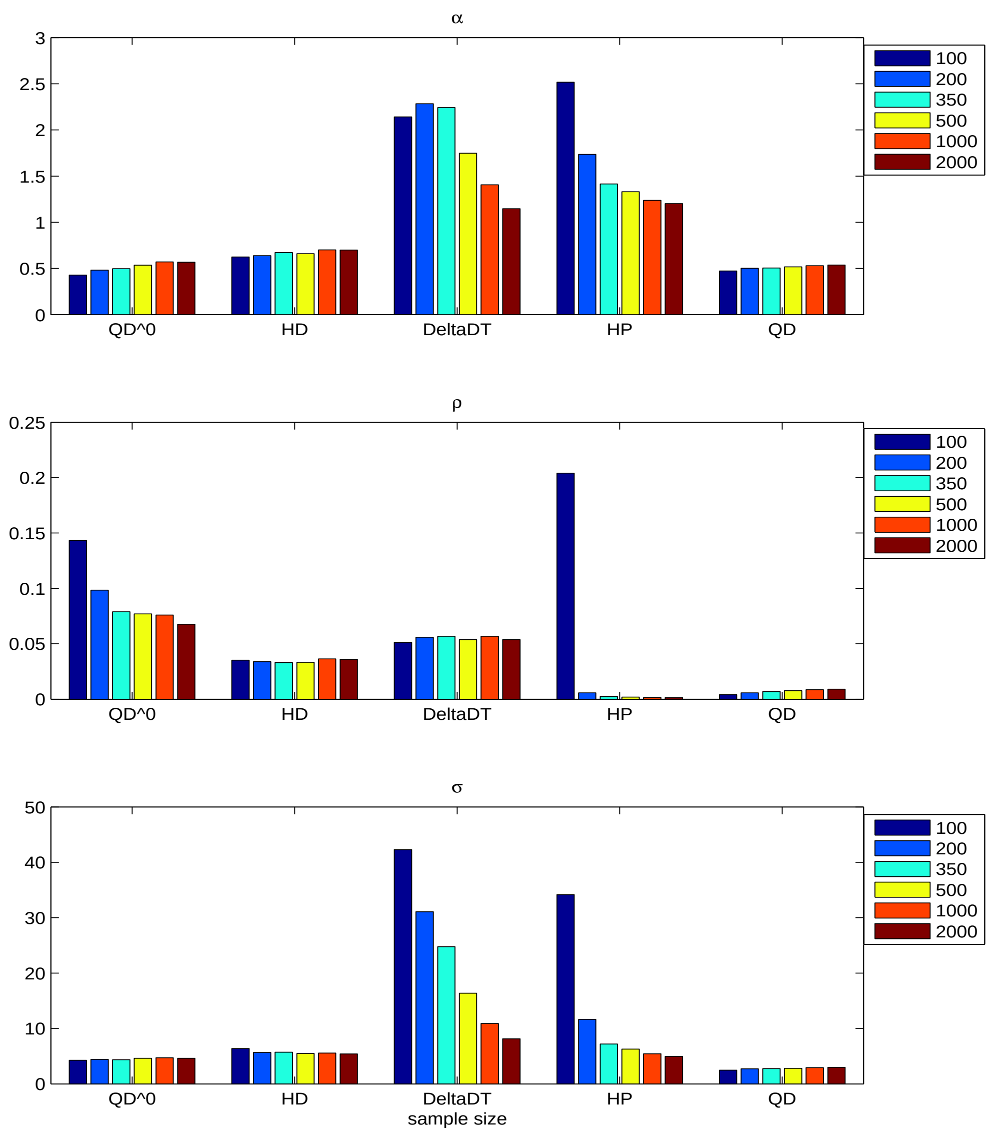

Figure 5: Base Case: $T\times$ MSE, $\rho = 1.0$

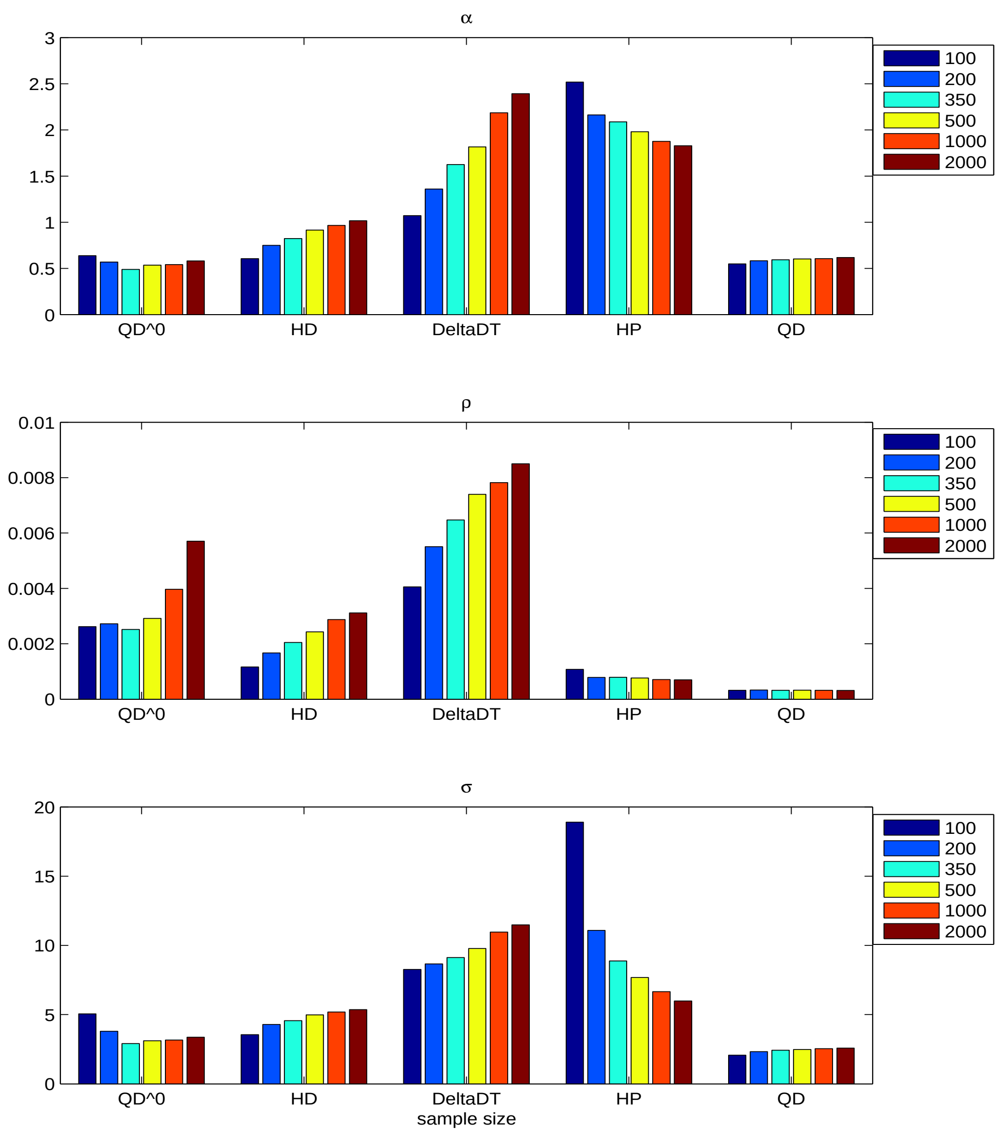

Figure 6: Other Models : $T\times$ MSE, $\rho = 0.95$

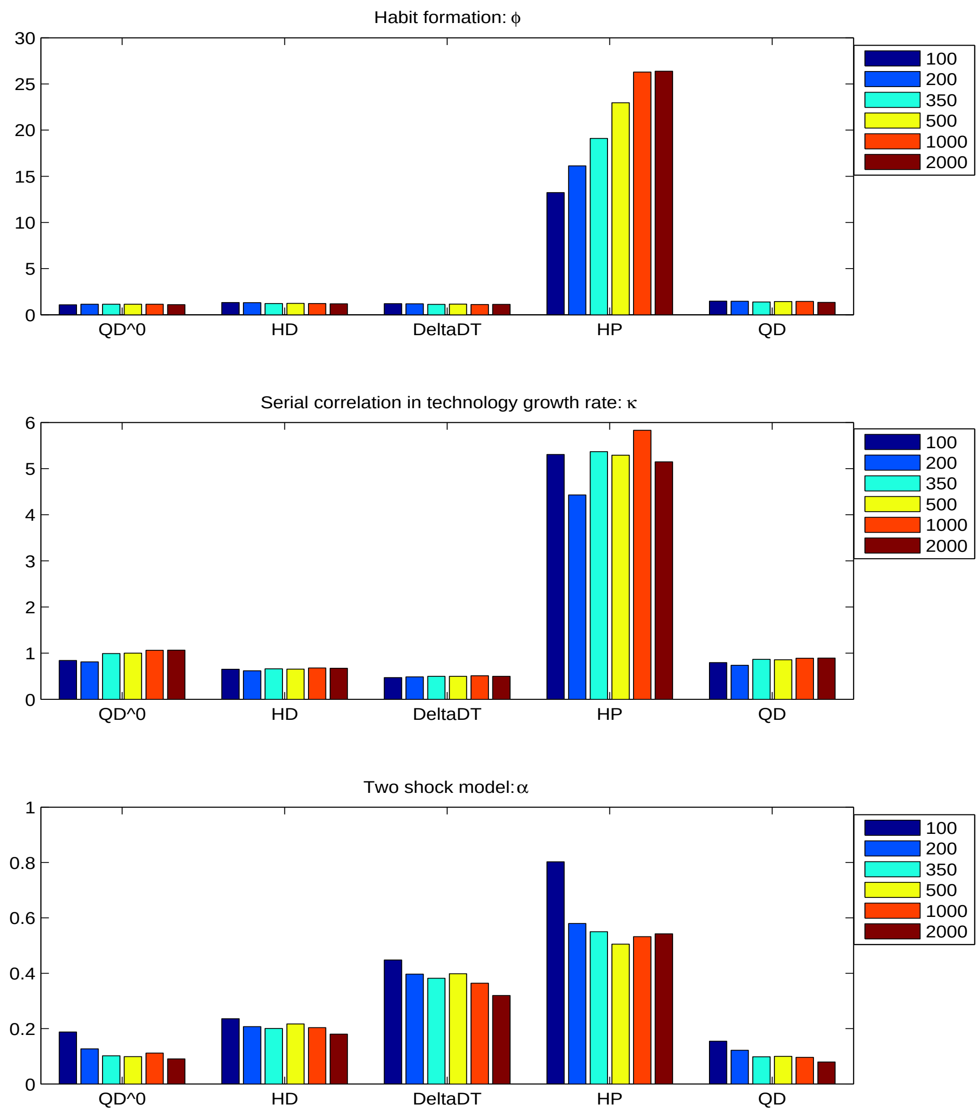

Figure 7: Other Models: $T\times$ MSE, $\rho = 1.0$