ESTIMATION OF DSGE MODELS WHEN THE DATA ARE PERSISTENT

Yuriy Gorodnichenko* Serena Ng†

July 2009

Abstract

Dynamic Stochastic General Equilibrium (DSGE) models are often solved and estimated under specific assumptions as to whether the exogenous variables are difference or trend stationary. However, even mild departures of the data generating process from these assumptions can severely bias the estimates of the model parameters. This paper proposes new estimators that do not require researchers to take a stand on whether shocks have permanent or transitory effects. These procedures have two key features. First, the same filter is applied to both the data and the model variables. Second, the filtered variables are stationary when evaluated at the true parameter vector. The estimators are approximately normally distributed not only when the shocks are mildly persistent, but also when they have near or exact unit roots. Simulations show that these robust estimators perform well especially when the shocks are highly persistent yet stationary. In such cases, linear detrending and first differencing are shown to yield biased or imprecise estimates.

*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 Brown University, the University of Michigan, the 2007 NBER Summer Institute, Princeton University, UC Berkeley, and the New York Area Macro Conference. We thank Marc Giannone,Tim Cogley, Anna Mikusheva, and two anonymous referees, the Associate Editor and the Editor 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.

Any attempt to estimate DSGE models must confront the fact that macroeconomic data are highly persistent. This fact often requires researchers to take a stand on the specification of the trends in DSGE models. Specifically, to take the model to the data, a researcher needs to use sample analogs of the deviations from steady states and, in doing so, must decide how to detrend the variables in the model and in the data. Table 1 is a non-exhaustive listing of how trends are treated in some notable papers.[1] Some studies assume stochastic trends for the model and use first differenced data in estimation. A number of studies specify deterministic trends for the model and use linearly detrended data in estimation. Studies that apply the Hodrick-Prescott (HP) filter to the data differ in what trends are specified for the model. Some assume simple linear trends, while others assume unit root processes. Table 1 clearly demonstrates that a variety of trends have been specified for the model and a variety of detrending methods have been used in estimation.

The problem for researchers is that it is not easy to ascertain whether highly persistent data are trend stationary or difference stationary in finite samples. While many have studied the implications for estimation and inference of inappropriate detrending in linear models,[2] much less is known about the effects of detrending in estimation of non-linear models. From simulation evidence of Doorn (2006) for an inventory model, it seems that HP filtering can significantly bias the estimated dynamic parameters. As well, while the local-to-unit framework is available to help researchers understand the properties of the estimated autoregressive root when the data are strongly persistent, it is unclear to what extent the framework can be used in non-linear estimation even in the single equation case. What makes estimation of DSGE models distinct is that they consist of a system of equations and

[1] As of June 2009, these papers were cited almost 2500 times at the Web of Science (former Social Science Citation Index) and almost 8000 times at Google Scholar.

[2] For example, Nelson and Kang (1981) showed that linear detrending a unit root process can generate spurious cycles. Cogley and Nason (1995a) found that improper filtering can alter the persistence and the volatility of the series while spurious correlations in the filtered data was documented in Harvey and Jaeger (1993). Singleton (1988) and Christiano and den Haan (1996) discussed how inappropriate filtering can affect estimation and inference in linear models.

misspecification in one equation can affect estimates in other equations.

This paper develops robust estimation procedures that do not require researchers to take a stand on whether shocks in the model have an exact or a near unit root, and yet obtain consistent estimates of the model parameters. All robust procedures have two characteristics. First, the same transformation (filter) is applied to both the data and the model variables. Second, the filtered variables are stationary when evaluated at the true parameter vector. The estimators have the classical properties of being $\sqrt{T}$ consistent and asymptotically normal for all values of the largest autoregressive root.

Our point of departure is that the not too uncommon practice of applying different filters to the model variables and the data can have undesirable consequences. Indeed, as we will show, estimates of parameters governing the propagation and amplification mechanisms in the model can be severely distorted when the trend specified for the model is not consistent with the one applied to the data. Accordingly, we insist on estimators that apply the same transformation to both the model and the data. This, however, may still lead to biased estimates if the filter does not remove the trends actually present in the data. Accordingly, we need to work with filters that can remove both deterministic and stochastic trends without the researcher taking a stand before solving and estimating the model. The idea of applying robust filters to both the model and the data is not new. Christiano and den Haan (1996) as well as Burnside (1998) applied the HP filter to both the model and the data, but they had to resort to simulation estimation to get around the large state vector that the HP filter induces. The filters we consider have the same desirable feature as the HP in that they adapt to the trends in the data. However, they can be implemented with simple modifications to the state space system while keeping the dimension of the state vector small. Specifically, we consider four transformations: (i) quasi-differencing, (ii) unconstrained first differencing, (iii) hybrid differencing, and (iv) the HP filter. All filters can be used in GMM estimation but not every method can be implemented in the likelihood framework. Importantly, one can use standard asymptotic inference as the finite sample distribution of the estimators are well approximated by the normal distribution not only when the large autoregressive root is far from one, but also when it is near or on the unit circle. The procedures can be applied to DSGE models whose solution can be shown to exist and is unique, and can be solved using variations of the method discussed in Blanchard and Kahn (1980) and Sims (2002).

As discussed in Canova and Sala (2009), DSGE models are susceptible to identification failure, in which case, consistent estimation of parameters is not possible irrespective of the treatment of trends. In view of this consideration and to fix ideas, we use a basic stochastic growth model whose properties are well understood. The model, which will be presented in Section 2, will also be used to perform baseline simulation experiments. The new estimators are presented in Sections 3 and 4. Sections 5 and 6 use simulations to show that the robust approaches perform well especially when the shocks are highly persistent yet stationary. These results also hold up in larger models though some filters are more sensitive to the number of shocks than others. In contrast, linear detrending and first differencing often lead to severely biased estimates. Implementation issues and related work in the literature are discussed in Section 7. Section 8 concludes.

2 Preliminaries

Consider the one sector stochastic growth model. The problem facing the central planner is:

$$\max E_t \sum_{t=0}^{\infty} \beta^t \vartheta(C_t, L_t)$$

subject to

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

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, e_t^z is an innovation in technology. Note that we allow ρ_z to be on the unit circle. The first order conditions are:

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

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. Given our assumptions, labor L_t is stationary for all $\rho_z \leq 1$ and thus $l_t^* = 0$. Collect the observed model variables into the vector $m_t = (c_t, k_t, l_t)$ and denote the trend component of the model variables by $m_t^* = (c_t^*, k_t^*, l_t^*)$. In general, how we define m_t^*, how we linearize the model, and how we estimate the parameters will depend on whether $\rho_z < 1$ or $\rho_z = 1$.

When $\rho_z < 1$, we have $c_t^* = k_t^* = \bar{g}t$, and thus $m_t^* = (\bar{g}t, \bar{g}t, 0)$. The detrended variables in the model are $\widehat{m}_t = (\widehat{c}_t, \widehat{k}_t, \widehat{l}_t) = (c_t - \bar{g}t, k_t - \bar{g}t, l_t) = m_t - m_t^*$. 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

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

Solving the system of expectation equations yields the reduced form

$$\begin{aligned} \widehat{m}_t &= \Pi\widehat{m}_{t-1}+Bu_t^z \qquad (1) \\ u_t^z &= \rho_z u_{t-1}^z + e_t^z. \end{aligned}$$

We assume that all roots of Π are strictly less than one, so that non-stationarity can only arise because ρ_z is on the unit circle. Note that when $\rho_z = 1$, the model needs to be linearized and solved with $m_t^* = (u_t + \bar{g}t, u_t + \bar{g}t, 0) = (z_t, z_t, 0)$. Despite the fact that the permanent shock u_t^z is now part of m_t^*, (1) is still the reduced form representation for *levels* of the linearly detrended variables. In other words, the reduced form representation for $\widehat{m}_t$ is continuous in ρ_z even though how we arrive at this representation will depend on ρ_z. Hence, without loss of generality, we will always use the representation (1) in subsequent discussions for all values of ρ_z.

Note that by definition, $\widehat{m}_t$ is the linearly detrended component of the model variables m_t. In other words, $\widehat{m}_t$ is a model concept. Hereafter, we let d_t denote the data analog of m_t. For the stochastic growth model, $d_t = (c_t, k_t, l_t)$ are the data collected for the empirical exercise. Let $\widehat{d}_t$ be obtained by removing deterministic trends from d_t. Then $\widehat{d}_t$ is the data analog of $\widehat{m}_t$.

3 Robust Estimators

In this section, we present robust methods that do not require the researcher to take a stand on the properties of trends in the data. We use the stochastic neoclassical growth model to illustrate the intuition behind the proposed methods.

Many estimators have been used to estimate DSGE models.[3] Our focus will be a method of moments (MM) estimator that minimizes the distance between the second moments of data and the second moments implied by the model, as in Christiano and den Haan (1996), Christiano and Eichenbaum (1992). Adaptation to likelihood based estimation will be discussed in Section 6. Let Θ denote the unknown structural parameters of the model and partition $\Theta = (\Theta^-, \rho_z)$. The generical MM estimator can be summarized as follows:

[3] Likelihood and Bayesian based methods, (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 (2007) provides a review of these methods.

1: Apply a filter (if necessary) to d_t and compute $\widehat{\Omega}^d(j)$, the estimated covariance matrix of the *filtered* series at lag j. Collect the data moments in the vector

$$\widehat{\omega}^d = (vech(\widehat{\Omega}^d(0))' \;\; vec(\widehat{\Omega}^d(1))' \; \ldots \; vec(\widehat{\Omega}^d(M))')'.$$

2: Solve the rational expectations model for a guess of Θ. Compute $\Omega^m(j)$, the model implied autocovariances of the *filtered* $\widehat{m}_t$ analytically or by simulation. Collect the model moments in the vector

$$\omega^m = (vech(\Omega^m(0))' \;\; vec(\Omega^m(1))' \; \ldots \; vec(\Omega^m(M))')'.$$

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

The choice of moments in MM can be important for identification (see e.g. Canova and Sala (2009)). We use the unconditional autocovariances but matching impulse responses can also be considered. Although MM is somewhat less widely used than maximum likelihood estimators in the DSGE literature, it does not require parametric specification of the error processes and it is easy to implement. The more important reason for using MM is practical as it can be used with many popular filters. We will return to this point subsequently.

The statistical properties of $\widehat{\Theta}$ will depend on ρ_z and the filters used. The robust approaches we consider always apply the same filter to the model and the data so that the filtered variables are stationary when evaluated at the true parameter of ρ_z, which can be one or close to one. We now consider four filters and then explore which of these have better finite samples properties. Properties of estimators that do not have these features will also be compared.

3.1 The QD Estimator

Let $\Delta^{\rho_z} = 1 - \rho_z L$ be the quasi-differencing (QD) operator and let $\Delta^{\rho_z}\widehat{m}_t = (1-\rho_z L)\widehat{m}_t$. Multiplying both sides of (1) by Δ^{ρ_z} and using $u_t^z = \rho_z u_{t-1}^z + e_t^z$ gives

$$\Delta^{\rho_z}\widehat{m}_t = \Pi\Delta^{\rho_z}\widehat{m}_{t-1} + Be_t^z. \tag{2}$$

Note that the error term in the quasi-differenced model is an i.i.d. innovation. As $\Delta^{\rho_z}\widehat{m}_t$ is stationary for all $\rho_z \leq 1$, its moments are well defined. In contrast, the moments of $\widehat{m}_t$ are not well defined when $\rho_z = 1$. This motivates estimation of Θ as follows:

1: Initialize ρ_z.

2: Quasi-difference $\widehat{d}_t$ with ρ_z to obtain $\Delta^{\rho_z}\widehat{d}_t$. Compute $\widehat{\Omega}^d_{\Delta^{\rho_z}}(j) = cov(\Delta^{\rho_z}\widehat{d}_t, \Delta^{\rho_z}\widehat{d}_{t-j})$, the sample autocovariance matrix of the quasi-differenced data at lag $j = 0, \dots, M$. Define

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

and let $\widehat{\omega}^d_{\Delta^{\rho_z}} = (vec(\widehat{\Upsilon}^d_{\Delta^{\rho_z}}(1))', \ \dots, \ vec(\widehat{\Upsilon}^d_{\Delta^{\rho_z}}(M))')'$;

3: For a given ρ_z and Θ^-, solve for the reduced form (1). Apply Δ^{ρ_z} to $\widehat{m}_t$ and compute $\Omega^m_{\Delta^{\rho_z}}(j), j = 1, \dots, M$, the model implied autocovariance matrices of the quasi-differenced variables. Let

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

Define $\omega^m_{\Delta^{\rho_z}}(\rho_z) = (vec(\Upsilon^m_{\Delta^{\rho_z}}(1))', \ \dots, \ vec(\Upsilon^m_{\Delta^{\rho_z}}(M))')'$;

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

The QD estimator is based on the difference between the model and the sample autocovariances of the filtered variables, normalized by the respective variance matrix $\Omega_{\Delta^{\rho_z}}(0)$. The QD differs from a standard covariance estimator 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 data quasi-differenced at ρ_z. As ρ_z and Θ^- are estimated simultaneously, the filter is data dependent rather than fixed. The crucial feature is that the quasi-transformed data are stationary when evaluated at the true ρ_z, which subsequently permits application of a central limit theorem. The normalization of the lagged autocovariances by the variance amounts to using the moments

$$cov(\Delta^{\rho_z}\widehat{d}_t, \ \Delta^{\rho_z}\widehat{d}_t - \Delta^{\rho_z}\widehat{d}_{t-j}) - cov(\Delta^{\rho_z}\widehat{m}_t, \ \Delta^{\rho_z}\widehat{m}_t - \Delta^{\rho_z}\widehat{m}_{t-j})$$

for estimation. The j-th difference of $\Delta^{\rho_z}\widehat{d}_t$ is always stationary and ensures that the asymptotic distribution is well behaved. Finally, observe that since we solve the model in levels and use the transformed variables only to compute moments, we preserve all equilibrium relationships between variables.

3.2 The FD Estimator

If $\Delta^{\rho_z}\widehat{m}_t$ is stationary when $\rho_z \leq 1$, the data vector is also stationary when quasi-differenced at $\rho_z = 1$. Denote the first-differencing (FD) operator by $\Delta = 1 - L$ and consider the following estimation procedure:

1: Compute $\widehat{\Omega}^d_{\Delta}(j) = cov(\Delta\widehat{d}_t, \Delta\widehat{d}_{t-j})$, the sample autocovariance matrix of the first differenced data at lag $j = 1, \dots, M$. Define $\widehat{\omega}^d_{\Delta} = (vech(\widehat{\Omega}^d_{\Delta}(1))', \ \dots, \ vec(\widehat{\Omega}^d_{\Delta}(M))')'$;

2: For a given Θ, solve for the reduced form (1). Compute $\Omega^m_\Delta(j)$, the model implied autocovariance matrices of the first-differenced variables $\Delta\widehat{m}_t$. Define $\omega^m_\Delta = (vech(\Omega^m_\Delta(0))',\ \ldots,\ vec(\Omega^m_\Delta(M))')'$;

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

To be clear, we compute autocovariances for first difference of the data and the model variables, but ρ_z is a free parameter which we estimate. Note that the QD and FD estimators are equivalent when $\rho_z = 1$. The key difference between FD and QD is that FD is a fixed filter while the QD is a data dependent filter.

3.3 The Hybrid Estimator

One drawback of the FD estimator is that when ρ_z is far from 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 $\widehat{\Omega}^d_{\Delta^{\rho_z}}(j)$ is quadratic in ρ_z. As will be explained below, this is why we normalize $\widehat{\Omega}^d_{\Delta^{\rho_z}}(j)$ by $\widehat{\Omega}^d_{\Delta^{\rho_z}}(0)$. These considerations suggest a hybrid estimator:

1: Transform the observed data to obtain $\Delta^{\rho_z}\widehat{d}_t$ (as in QD) and $\Delta\widehat{d}_t$ (as in FD).

2: Compute $\widehat{\Omega}^d_{QD,\Delta}(j) = cov(\Delta^{\rho_z}\widehat{d}_t, \Delta\widehat{d}_{t-j})$, the covariance between $\Delta^{\rho_z}\widehat{d}_t$ and $\Delta\widehat{d}_{t-j}$. Define $\widehat{\omega}^d_{QD,\Delta} = (vec(\widehat{\Omega}^d_{QD,\Delta}(0))',\ \ldots,\ vec(\widehat{\Omega}^d_{QD,\Delta}(M))')'$;

3: For a given Θ, solve for the reduced form (1), and compute the model implied autocovariances between the quasi-differenced and the first differenced variables. Define

$$\omega^m_{QD,\Delta} = (vec(\Omega^m_{QD,\Delta}(0))', \ldots, vec(\Omega^m_{QD,\Delta}(M))')';$$

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

Notice that $\widehat{\Omega}^d_{QD,\Delta}(j)$ is now linear in ρ_z, unlike $\widehat{\Omega}^d_{\Delta^{\rho_z}}(j)$. We denote this estimator with HD (hybrid differencing).

3.4 The HP Estimator

Some linear filters such as the HP and the bandpass can also remove deterministic and stochastic trends, see Baxter and King (1999) and King and Rebelo (1993). In this paper, we will focus only on the HP filter, which is heavily used in empirical analysis. An HP detrended series is defined as

$$HP(L)d_t \quad = \quad \frac{\lambda(1-L)^2(1-L^{-1})^2}{1+\lambda(1-L)^2(1-L^{-1})^2}d_t.$$

Given this transformation, the procedure can be summarized as follows:

1: Compute the autocovariance matrices of the HP-filtered data $\widehat{\Omega}^d_{HP}(0),\ \ldots,\widehat{\Omega}^d_{HP}(M)$. Define $\widehat{\omega}^d_{HP} = (vech(\widehat{\Omega}^d_{HP}(0))',\ \ldots,\ vec(\widehat{\Omega}^d_{HP}(M))')'$;

2: For a given guess of Θ, solve for the reduced form (1), and compute $\Omega^m(j)$, the autocovariances of $\widehat{m}_t$. Apply the Fourier transform to obtain the spectrum for $\widehat{m}_t$ at frequencies $2\pi s/T,\ s = 0, \ldots, T-1$. 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 $\omega^m_{HP} = (vech(\Omega^m_{HP}(0))',\ \ldots,\ vec(\Omega^m_{HP}(M))')'$.

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

This approach is similar to Burnside (1998) who also applies the HP filter to both the model and the data series, but he uses simulations to compute model-implied moments. Like the FD, the variables in the model and the data will be stationary without ρ_z entering the filter. By construction of the HP filter, both the filtered data and the filtered model variables are stationary for all $\rho_z \leq 1$. Note that HP filtering involves (implicitly or explicitly) estimation of many more autocovariances than the other estimators considered above. Burnside (1998) reports that although the HP filter removes variation potentially informative for estimation of the structural parameters, the HP filtered model and data series have sufficient variability to discriminate competing theories of business cycles.

4 Properties of the Estimators

Let $\widehat{\omega}^d_j$ generically denote the $j-th$ sample moments of the filtered variables while $\omega^m_j(\Theta)$ denote the model moment based on the same filter. Define

$$\bar{g}_j(\Theta) = \widehat{\omega}^d_j - \omega^m_j(\Theta)$$

and let $\bar{g}(\Theta) = (\bar{g}_0(\Theta), \bar{g}_1(\Theta), \ldots, \bar{g}_M(\Theta))$. Then the MM estimator is $\widehat{\Theta} = \min_\Theta ||\bar{g}(\Theta)||$ is a non-linear GMM estimator using an identity weighting matrix. This sub-optimal weighting matrix is used because when there are fewer shocks than variables in the system, stochastic singularity will induce collinearity in the variables resulting in a matrix of covariances that would be singular. Even if there are as many shocks as endogenous variables, Abowd and Card (1989), Altonji and Segal (1996) and others find that an identity matrix performs better than the optimal weighting matrix in the context of estimating covariance structures. The optimal weighting matrix, which contains high order moments, tends to correlate with the moments and this correlation undermines the performance of the estimator.

Let $\bar{G}(\Theta)$ be the matrix of derivatives of $\bar{g}(\Theta)$ with respect to Θ. In standard covariance structure estimation, the parameters enter the model moments $\omega^m(\Theta)$ but not the sample $\widehat{\omega}^d$, so that if $\widehat{\omega}^d$ are moments of stationary variables, then under regularity conditions such as stated in Newey and McFadden (1994), we have the conventional result that the estimators are consistent and

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

where $A = (G_0' G_0)^{-1} G_0'$ and $\sqrt{T}\bar{g}(\Theta_0) \xrightarrow{d} N(0, S)$, and G_0 is the probability limit of $\bar{G}(\Theta)$ evaluated at $\Theta = \Theta_0$. This distribution theory applies to the FD and the HP because these two filters do not depend on unknown parameters and the filtered variables are always stationary. For the HD estimator, $\widehat{\omega}^d$ depends on ρ_z but its first derivative does not, so that a quadratic expansion of the objective function can still be used to derive the asymptotic distribution of the estimator. Although $\bar{G}(\Theta)$ for the HD has a random limit when $\rho_z = 1$, $\widehat{\omega}^d$ is a vector of covariances of stationary variables when evaluated at the true value of ρ_z. Thus, the 'standardized' estimator (or the t statistic) remains asymptotically normal.

To understand the properties of the QD estimator, we need to explain why we normalize the lagged autocovariances by the variance. Suppose we had used $\widetilde{g}_j(\Theta) = \widehat{\Omega}^d_{\Delta^{\rho_z}}(j) - \Omega^m_{\Delta^{\rho_z}}(j)$ instead of $\bar{g}_j(\Theta) = \widehat{\omega}^d_{\Delta^{\rho_z}}(j) - \omega^m_{\Delta^{\rho_z}}(j)$ where $\omega^d_{\Delta^{\rho_z}}(j) = \widehat{\Omega}^d_{\Delta^{\rho_z}}(j) - \widehat{\Omega}^d_{\Delta^{\rho_z}}(0)$, and $\omega^m_{\Delta^{\rho_z}}$ is likewise defined. Minimizing $||\widetilde{g}(\Theta)||$ over Θ yields an estimator that we can call QD^0. The problem here is that $\widehat{\Omega}^d_{\Delta^{\rho_z}}(j)$ is a cross-product of data quasi-differenced at ρ_z, and is thus quadratic in ρ_z. The quadratic expansion of $||\widetilde{g}(\Theta)||$ around Θ_0 contains terms that are not negligible when ρ_z is one. As such, the sample objective function cannot be shown to converge uniformly to the population objective function. In fact, we show in Gorodnichenko et al. (2009) in a simpler setting that the QD^0 estimator for ρ_z is consistent but it has a convergence rate of $T^{3/4}$ and is not asymptotically normal. The QD estimator is motivated by the fact that the offending term in the quadratic expansion of $\widehat{\Omega}^d_{\Delta^{\rho_z}}(j)$ is collinear with $\widehat{\Omega}^d_{\Delta^{\rho_z}}(0)$ when $\rho_z = 1$.

Proposition 1 *Consider a DSGE model whose reduced form is given by (1) and all roots of* Π *less than one. Let* Θ *be the unknown parameters of the model and let* $\widehat{\Theta}^{QD}$ *be the QD estimator of* Θ*. Then* $\sqrt{T}(\widehat{\Theta}^{QD} - \Theta_0) \xrightarrow{d} N(0, Avar(\widehat{\Theta}^{QD}))$.

A sketch of the argument is given in the Appendix for the baseline model whose closed-form solution is known. By subtracting the variance from each lagged autocovariance, the quadratic terms in the expansion of the objective function are asymptotically negligible. This leads to the rather unexpected property that $\widehat{\Theta}$ is asymptotically normal even when $\rho_z = 1$. From a practical perspective, the primary advantage of the robust estimators is that when properly studentized, the estimators are normally distributed whether $\rho_z < 1$ or $\rho_z = 1$, which greatly facilitates inference. Since all estimators are consistent and asymptotically normal, it remains to consider which estimator is more efficient in finite samples.

4.1 Related Literature

Our approach is related to other methods considered in the literature. Cogley (2001) 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 (and complementary) 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 often 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 exploit different properties of the model in estimation and hence can be used in a broader array of situations.

Fukac and Pagan (2006) consider how the treatment of trends might affect estimation of DSGE models, but their analysis is confined to a single equation framework. They propose to use the Beveridge-Nelson decomposition to estimate and remove the permanent components in the data. Apart from the fact that the permanent component in the Beveridge-Nelson decomposition may be different from actual trend and is derived under some strong assumptions, our approach is a one-step procedure that can handle multiple I(1) shocks as we discussed above.

Canova (2008) explicitly treats the latent trends as unobserved components and estimates the trends and cycles directly. While this allows the data to select the trend endogenously, the procedure can be imprecise when the random walk component is small. Canova and Ferroni (2008) considered many filters and treat each as the true cyclical component measured with error. They are primarily concerned with the consequences of data filtering taking the model specification as given. We take the view that the trends specified for the model should be consistent with the facts that we sought to explain. As such, it cannot be taken as given.

5 Simulations: Baseline Model

In this section we use the stochastic growth model to conduct Monte Carlo experiments. We generate data with either deterministic trends ($\rho_z < 1$) or stochastic trends ($\rho_z = 1$) using the model equations for $\widehat{m}_t$. The model variables are then rescaled back to level form and treated as observed data $d_t = (c_t, k_t, y_t, l_t)$, which we take as given in estimation. We use the variance and first order autocovariance (including cross variances) of the four variables as moments. Thus, $M = 1$. We also experimented with alternative choices of observed variables, such as excluding the capital stock series, and we found very similar results.

We estimate $\Theta = (\alpha, \rho, \sigma)$ and treat parameters (β, δ, θ) as known.[4] We estimate only a handful of parameters because we want to decouple the issues related to the treatment of trends from the identification issues that might arise. 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 parameter ρ_z to take values (0.95, 0.99, 1). We have only one shock in this baseline model. Thus, we set the standard deviation of e_t^z to $\sigma_z = 1$ without loss of generality. We perform 2,000 replications for each choice of parameter values. In each replication, we create series with T=200 observations. Other sample sizes are also considered.

In all simulations and for all estimators, we set the starting values in optimization routines equal to the true parameter values. The model is solved using the Anderson and Moore (1985) algorithm. We allow 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 only keep as solution those sets of estimates that are consistent with a unique rational expectations equilibrium.[5]

We report simulation results for the baseline neoclassical growth model in Table 2. The persistence of technology shocks is given in the left column. The first and second rows indicate which filter is applied to both the data and the model variables. Columns (1)-(4) report results for the four estimators. By and large, all four filters yield estimates which are very close to the true values. Notice that while ρ_z is always precisely estimated, the variance of the estimates varies substantially across filters. The QD estimates has the lowest standard deviation while the HP estimates are two to five

[4] The average growth rate $\bar{g}$ is estimated in the preliminary step when we project series on linear time trend.

[5] 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. This is distinct from covariance stationarity of the time series data, which 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.

times more variable than the QD. The HD is more precise than the FD but is less precise than the QD. This pattern is recurrent in all simulations.

Figure 1 shows the root mean squared error (RMSE) for different estimators and sample sizes. The QD estimator performs the best while the FD tends to have the largest RMSE in almost all cases. The performance of the HP estimator varies with sample size. In small samples, the HP tends to lead to large RMSE while in larger samples, the HP approaches the HD which is only slightly inferior to the QD.

Figures 2 through 4 present the kernel density of the normalized estimator (ie. $\sqrt{T}(\widehat{\Theta} - \Theta)$) for sample sizes of T=150 and 300. Results are also reported for T=2000 to study the asymptotic properties of the estimators. Approximate normality of $\widehat{\rho}_z$ when ρ_z is 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. Instead, all densities are bell-shaped and symmetric for all $\rho_z \leq 1$ with no apparent discontinuity as we increase ρ_z to one. The normal approximation is not perfect in small samples, suggesting that some size distortion will occur if we use the t statistic for inference. In results not reported, we construct t-statistics using Newey-West standard errors and find that the rejection rates tend to be greater than the nominal size for all estimators except the HP, which can be undersized. For example, the rejection rate of the QD estimator for the two-sided t-test of ρ_z at the true value of 1 is 0.055 when T=200 while for testing α at the true value of 0.33, the rejection rate is 0.21. This is larger than the nominal size of 0.05. As the sample size increases, the actual size gets closer (and eventually converges) to the nominal rates. For example, at T=1000 for QD, the two-sided t-test of $\rho_z = 1$ has a rejection rate of 0.05, while the t test for $\alpha = .33$ is 0.10. The QD and HD generally have better size than the FD and the HP. The finite sample size distortion seems to be a general problem with covariance structure estimators and not specific to the estimator we consider. Burnside and Eichenbaum (1996) reported similar results in covariance structure estimation with as many overidentifying restrictions as we have, and also using the Newey-West estimator of the variance of moments.

5.1 Variations to the Baseline Model

In response to the finding in Cogley and Nason (1995b) that the basic real business cycle model has weak internal propagation, researchers often augment the basic model to strengthen the propagation and to better fit the data at business cycle frequencies. One consideration is to introduce serial correlation in the growth rate of shocks to technology by assuming $u_t = (\rho_z + \kappa)u_{t-1} - \kappa\rho_z u_{t-2} + e_t^z$. This specification generates serial correlation of κ in the growth rate of technology when $\rho_z \approx 1$. Our baseline model corresponds to $\kappa = 0$. When we simulate data with $\kappa = 0$ and estimate κ freely, the

QD, HD, FD, and HP correctly find that $\kappa = 0$ (Table 3, Panel A).

Habit in consumption is another popular way to introduce greater persistence in business cycle models. Consider the utility function: $\vartheta(C_t, C_{t-1}, L_t) = \ln(C_t - \phi C_{t-1}) - \theta L_t$ where ϕ measures the degree of habit in consumption. We set $\phi = 0$ and estimate ϕ along with other parameters to investigate how the treatment of the trends might affect estimates of this internal propagation mechanism. The robust estimators again find $\widehat{\phi}$ to be numerically small and not statistically from zero for all values of ρ_z (Table 3, Panel B).

We also augment our baseline model with a preference shock Q_t such that the utility is $\vartheta(C_t, L_t) = \ln C_t - \theta L_t/Q_t$ where $q_t = \ln Q_t = \rho_q q_{t-1} + e_t^q$ and $e_t^q \sim iid(0, \sigma_q^2)$. In our simulations, we set $\rho_q = 0.8$ so that the preference shock is stationary. We let $\sigma_q = (0.5, 1.0, 1.5)$. To conserve space, we only report estimates for α in Table 3, Panel C. Consistent with the results thus far, the HP estimates have the largest variability although the difference with other estimators is not as large as it was in the baseline model. Note that as we increase σ_q, the difference across methods shrinks while the precision for all estimators improves.

A recurrent result is that the HP estimates have the largest variability. One possibility is that the HP filters out more low frequency variation than other filters, and the parameters ϕ and κ are identified from these frequencies (see also discussion in Burnside (1998)). Another possibility is that the HP implicitly uses estimated autocovariances at many more lags than other filters (recall that we apply the inverse Fourier transform to many autocovariances). This extensive use of sample autocovariances can also introduce variability to the estimator. In addition, we find that HP is much more computationally intensive than other robust estimators.

6 Non-Robust Estimators and a Model with Multiple Rigidities

In this section, we report results of the non-robust estimators applied to different models to illustrate how treatment of trends can lead to misleading conclusions about the propagating mechanism of shocks. We also compare the estimators using a model with many more endogenous variables.

6.1 Alternative Detrending Procedures

Up to this point, we have considered approaches where the same transformation is applied to the data and the model variables. Much has been written about the effects of filtering on business cycle facts. King and Rebelo (1993) and Canova (1998) showed that the HP filtered data are qualitatively different from the raw data. Canova (1998) showed that the stylized facts of business cycles are sensitive to the filter used to remove the trending components. Gregory and Smith (1996) used a calibrated business cycle model to investigate what type of trend can produce a cyclical component in the data that is similar to the cyclical component in the model. Although these authors did not estimate a DSGE model on filtered data, they hinted that the estimates of the structural parameters can be adversely affected by filtering.

We now investigate the consequences of using different and/or inappropriate filters by considering four combinations:

A) The autocovariances are computed for linearly detrended model and data series;

B) The autocovariances are computed for the first differenced model and data series with imposed $\rho_z = 1$;

C) The sample autocovariances are computed for HP filtered data but the model autocovariances are computed for the linearly detrended variables;

D) The sample autocovariances are computed for HP filtered data. The model autocovariances are computed for series normalized by the level of technology, i.e., $m_t - z_t$ where z_t is the level of technology.

Each combination has been used in the literature (see e.g. Table 1). (A) and (B) are aimed to show the effects of imposing incorrect assumptions about trends. (C) and (D) illustrate the consequences when different trends are applied to the model and the data. As a general observation, the starting values are very important for non-robust methods as the optimization routines can get stuck in local optima. With the robust estimators, the converged estimates do not change as we start the optimization from values other than the true parameters, though the search for global minimum was often long.

The results are reported in Table 2. For (A), which is reported in column (5), we see that when $\rho_z = 0.95$, the parameter estimates are slightly biased. As ρ_z increases, the estimates are strongly biased. This shows that when ρ_z is close to unity yet stationary, assuming trend stationarity still yields imprecise estimates. At $\rho_z = 1$, the mean of $\widehat{\rho}_z$ is 0.694 (instead of 1), the mean of $\widehat{\alpha}$ is approximately 0.905 (instead of 0.33), the mean of $\widehat{\sigma}_z$ is 19.8 (instead of 1). The case of $\rho_z \leq 1$ is empirically relevant because macroeconomic data are highly persistent and well approximated by unit root processes. Our results show that linear detrending of nearly integrated data in non-linear estimation can lead to biased estimates of the structural parameters. This resembles the univariate finding of Nelson and Kang (1981) that projecting a series with a unit root on time trend can lead to spurious cycles.

Turning to (B) in column (6) of Table 2, we find that while the estimates are fairly precise when ρ_z is indeed equal to one, as ρ_z departs from one, the estimates get increasingly biased. Hence imposing a stochastic trend when the data generating process is trend stationary can lead to seriously distorted estimates. Results for combination (C) are reported in column (7) of Table 2. The estimates of ρ_z are downward biased while $\widehat{\alpha}$ and $\widehat{\sigma}_z$ are upward biased. Taken at face value, these estimates suggest a significant role for capital as a mechanism for propagating shocks in the model.

Results for (D) are reported in column (8) of Table 2. Here, the estimates of α often hit the boundary of the permissible space while estimates of σ_z are close to zero. The reason is that when z_t has a unit root, shocks to $m_t - z_t$ are transitory. Thus, the endogenous variables such as consumption adjust quickly to the permanent technology shock. But the HP filtered data are serially correlated. Thus, the estimator is forced to produce parameter values that can generate a strong serial correlation in the model variables. Results for (C) and (D) are consistent with the findings of Cogley and Nason (1995a), King and Rebelo (1993) and Harvey and Jaeger (1993). These papers suggest 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 serial correlation of the filtered data.

Clearly, the large estimates of α will alert the researcher that the model is likely misspecified. Suppose the researcher allows for serially correlated shocks in technology growth by estimating κ freely. Panel A, Table 3 shows that the non-robust methods now 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 $\widehat{\kappa}$ strongly negative and statistically significant when the true value is zero. Suppose now the researcher modifies the model by allowing for habits in consumption. Evidently, the estimated habit formation parameter ϕ is sensitive to which non-robust estimator is used. In particular, (A) has a strong downward bias, while (B) produces a negative bias in $\widehat{\phi}$ when ρ_z departs from one. On the other hand, (C) and (D) have a strong upward bias. With either modification, 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 mislead the researcher to augment correctly specified models with spurious propagation mechanisms to match the moments of the data.

Results for the model with an additional labor supply shock are reported in Table 3, Panel C. The estimates continue to be biased although the biases tend to be smaller than in the baseline model with a single persistent shock. In general, a smaller ρ_z and a larger σ_q lead to smaller biases. In some cases, we find $\widehat{\sigma}_q > \widehat{\sigma}_z$, so that the researcher may be tempted to conclude that preference shocks have larger volatility than shocks to technology while the opposite is true.

6.2 The Smets and Wouters Model

Although the baseline model is an illuminating laboratory to evaluate how the estimators work, it is overly simplistic. We now consider the model of Smets and Wouters (2007) (henceforth SW). We use

SW's estimates for the post-1982 sample as the true values. We then simulate series of size T=150 and apply the estimators to the simulated series. To separate identification issues from issues related to the treatment of trends, we estimate only four parameters: persistence of technology shocks ρ_z whose true value varies across simulations, investment adjustment cost ϕ whose true value is 5.48, external habit formation in consumption λ whose true value is 0.71, and Calvo's probability of wage adjustment ξ_w whose true value is 0.73.

The results are reported in Table 4. All robust methods yield precise estimates of the parameters. Although the HP continues to be less precise, the difference with the other three robust estimators is smaller than in the baseline model. This is similar to what we observed when we compare the two-shock and one-shock neoclassical growth models. These differences between the baseline and the more complicated models can occur for several reasons. First, the SW model has six other structural shocks so that technology shocks explain only a fraction of variation in key macroeconomic variables. The HP estimator may simply need more shocks to identify the parameters. Second, the SW model imposes many cross equation restrictions on the handful of the parameters we estimate. These restrictions may improve the efficiency of some estimators more than others. Third, some estimators may be more sensitive to the size of the model than others. The general observation, however, is that our robust estimators perform reasonably well for all values of ρ_z in simple and more complex models.

In contrast, the non-robust estimators (A) through (D) have dramatic biases in all four parameters being estimated when (i) the filter used for the model and the data are different, when (ii) the assumed trends are different from trends in the data generating process, or when (iii) the data are stationary but highly persistent. Obviously, the impulse responses (and other analyses about the role of rigidities in amplification and propagation of shocks in business cycle models) based on these biased estimates of the structural parameters will be misleading. As an illustration, Figure 5 highlights the difference between the true response of key macroeconomic variables to a technology shock in the SW model and the responses based on parameter estimates from approaches (A) through (D). For instance, consider the response of consumption. Estimates from approaches (A) and (C) imply grossly understated responses. Estimates from approach (D) suggest a considerably more delayed consumption response than the true one. The consumption response implied by approach (B) is qualitatively similar to the true response, but the responses are noticeably different quantitatively especially when ρ_z is further away from one.

7 Extensions and Implementation Issues

7.1 Multiple shocks

The reduced form solution (1) can be easily generalized to other models and takes the form

$$\begin{aligned} \widehat{m}_t &= \Pi \widehat{m}_{t-1} + B u_t \\ u_t &= \rho u_{t-1} + S e_t, \end{aligned} \tag{3}$$

where u_t is now a vector of exogenous forcing variables, e_t is a vector of innovations in u_t, and the matrices Π, B, S, ρ are of conformable sizes.

Suppose there are J univariate shock processes, each characterized by

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

where some J^* 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. For example, if one knows that shocks to tastes dissipate quickly while technology shocks z_t are highly persistent, we can still use $(1 - \rho_z L)$ as Δ^ρ. Once the model is solved to arrive at (3), we can compute moments for the quasi-differences of $\widehat{m}_t$. Whether none, one, or more shocks are permanent, the autocovariances of the transformed variables are well defined.

7.2 Likelihood Estimation

As likelihood and Bayesian estimation is commonly used in the DSGE literature, one may wonder how the ideas considered in this paper can be implemented in likelihood based estimation. Conceptually, if we can write the model in a state space form, we can specify the likelihood which makes MLE and Bayesian estimation possible. This involves using the measurement equations to establish a strict correspondence between the detrended series in the model and in the data.

As an example, consider the FD estimator for the generalized model (3). We can define the measurement equation as

$$x_t = H s_t = \begin{bmatrix} \Psi & -\Psi & 0 \end{bmatrix} s_t, \tag{4}$$

where x_t is the vector of filtered variable, Ψ is the selection matrix, and $s_t' = (\widehat{m}_t, \widehat{m}_{t-1}, u_t)$ is the state vector. The corresponding transition equation is

$$\begin{bmatrix} \widehat{m}_t \\ \widehat{m}_{t-1} \\ u_t \end{bmatrix} = \begin{bmatrix} \Pi & 0 & B\rho \\ I & 0 & 0 \\ 0 & 0 & \rho \end{bmatrix} \begin{bmatrix} \widehat{m}_{t-1} \\ \widehat{m}_{t-2} \\ u_{t-1} \end{bmatrix} + \begin{bmatrix} BS \\ 0 \\ S \end{bmatrix} e_t$$

or

$$s_t = \Pi^* s_{t-1} + B^* e_t \tag{5}$$

with $e_t \sim i.i.d.(0, \Sigma)$. The measured variable x_t is stationary irrespective of whether $\widehat{m}_t$ has stochastic or deterministic trends. For the QD_0 estimator, $H = \begin{bmatrix} \Psi & -\rho\Psi & 0 \end{bmatrix}$. As with all quasi-differencing estimators, the treatment of initial condition is important especially when there is strong persistence. In simulations, we condition on the first observation being fixed and find that the MLE version of the FD gives precise estimates, but the t statistics are less well approximated by the normal distribution compared to MM-FD (see Figure 6).

For the other three estimators, the extension to MLE is either not possible or not practical. For MLE-HP, we would need to write out the entire data density of the HP filtered data, and the Jacobian transformation from the unfiltered to filtered data involves an infinite dimensional matrix. For the QD estimator, recall that we normalized the autocovariances by the variance in constructing the QD. By analogy, MLE-QD would require modifying the score vector. Although such modification is possible in theory, it is beyond the scope of this paper. We leave this promising idea for future research. For the HD estimator, the MLE implementation is cumbersome because HD exploits covariances of variables computed with different filters. The difference between the MM and MLE really boils down to a choice of moments, and the MM is more straightforward to implement.[6]

7.3 Computation

Computing Filtered Autocovariance Moments of the filtered model variables can be computed analytically or by using simulations. We use the analytical moments whenever possible since it tends to be much faster than simulations and it does not have simulation errors. Although there are a variety of methods for analytical calculations, methods that exploit measurement equations are computationally attractive especially in large models. Combining the measurement equation $x_t = Hs_t$ and the state equation $s_t = \Pi^* s_{t-1} + B^* e_t$, we have

$$\begin{bmatrix} x_t \\ s_t \end{bmatrix} = \begin{pmatrix} 0 & H\Pi^* \\ 0 & \Pi^* \end{pmatrix} \begin{bmatrix} x_{t-1} \\ s_{t-1} \end{bmatrix} + \begin{pmatrix} HB^* \\ B^* \end{pmatrix} e_t.$$

Let $w_t' = (x_t', s_t')$. We have

$$w_t = D_0 w_{t-1} + D_1 e_t.$$

[6]One practical drawback of GMM is perhaps that when many parameters have to be estimated, the objective function can be ill behaved which frustrates convergence. However, Chernozhukov and Hong (2003) suggest a novel approach by which GMM can take advantage of MCMC methods. Coibion and Gorodnichenko (2008) use Chernozhukov-Hong's MCMC method based on GMM to estimate a relatively large DSGE model.

The variance matrix $\Omega_w(0) = E(w_t w_t')$ can be computed by iterating the equation

$$\Omega_w^{(i)}(0) = D_0 \Omega_w^{(i-1)}(0) D_0' + D_1 \Sigma D_1' \tag{6}$$

until convergence. The autocovariance matrices can then be computed as $\Omega_w(j) = D_0^k \Omega_w(0)$. Since we are only interested in computing the moments of variables in the measurement vector x_t, we iterate equation (6) until the block that corresponds to x_t converges, i.e. $\|\Omega_x^{(i)}(0) - \Omega_x^{(i-1)}(0)\| < \epsilon$.[7]

Calculating moments of HP-filtered model series The HP filtered series can alternatively be obtained as follows:

$$HP(L)d_t = HP^+(L)\Delta d_t = \frac{\lambda(1-L)(1-L^{-1})^2}{1+\lambda(1-L)^2(1-L^{-1})^2}\Delta d_t.$$

In practice, using $HP^+(L)$ and the autocovariances for Δd_t and $\Delta\widehat{m}_t$ tend to give more stable results when ρ_z is close to one. It is possible to speed up estimation based on HP filtered series by using a smaller number of leads and lags. This modification however would deteriorate the HP's approximation to the desired filter removing low frequencies.[8]

Treatment of stationary variables Recall that in the stochastic growth model, $m_t^* = (\bar{g}t, \bar{g}t, 0)$ when $|\rho_z| < 1$ and $m_t^* = (u_t + \bar{g}t, u_t + \bar{g}t, 0)$ when $|\rho_z| = 1$, where the third component of m_t^* is the trend for labor supply, l_t. Since l_t has no deterministic or stochastic trend component, the autocovariances are computed for l_t and not $\widehat{l}_t$, though the results do not change materially if we had filtered these series as well. In general, if the j-th component of m_t^* is zero, it is understood that the autocovariances are computed for the level of the variable both in the model and in the data. An alternative is to deal with these non-trending variables through the measurement equation. Then some variables can be quasi-differenced or first-differenced, while others require no transformation.

8 Concluding Remarks

A realistic situation encountered with estimation of DSGE model is that (a) the data are trending; (b) deviations from the trend are persistent; (c) the researcher does not know whether the data generating process is difference or trend stationary. We document that the treatment of trends can significantly affect the parameter estimates of DSGE models and propose several robust approaches that produce

[7]Since parts of w_t may be exploding, we found that it useful to limit the size of elements in the covariance matrix $\Omega_w^{(i)}(0)$ by a fixed large number. This introduces an error in the calculated moments of x_t but we found that this error is negligible in practice.

[8]We also experimented with a simulation procedure. For each Θ, we use the model to generate $j = 1, \ldots, R$ samples of size T. For each j we computed moments. Then we average the moments over j and use this average for ω_{HP}^m. This procedure is computationally more intensive and the results are similar to the one considered here.

precise estimates without the researcher having to take a stand of trend specification. The key is to apply the same filter to the data and the model variables to yield well-defined moments for the estimation of the structural parameters. We consider several filters that can be used in methods of moments estimation, and the estimators have approximately normal finite sample distributions. Undoubtedly, the estimators require further scrutiny and can be improved in various dimensions.[9] Our analysis is a first step in the sparse literature on non-linear estimation when the data are highly persistent.

References

Abowd, J. M. and Card, D. 1989, On the Covariance Structure of Earnings and Hours Changes, *Econometrica* **57(2)**, 411–445.

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

Altonji, J. and Segal, L. 1996, Small-sample bias in GMM estimation of covariance structures, *Journal of Business and Economic Statistics* **14(3)**, 353–366.

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**, 575593.

Blanchard, O. and Kahn, C. 1980, The Solution of Linear Difference Models under Rational Expectations, *Econometrica* **48**, 1305–1313.

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

Burnside, C. 1998, Detrending and Business Cycle Facts: a Comment, *Journal of Monetary Economics* **41**, 513–532.

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.

Canova, F. 1998, Detrending and Business Cycle Facts, *Journal of Monetary Economics* **41**, 475–512.

Canova, F. 2008, Estimating DSGE models with Unfilterred Data, *mimeo.*

[9]For example, one can use the bootstrap developed for covariance structures in Horowitz (1998) to correct for small-sample biases because the asymptotic distribution of the t-statistic is normal. One might also consider a model-based weighting matrix computed as follows. Given a set of parameters, simulate the model to generate series and compute the moments. By repeating this many times, we can compute covariance of simulated moments rather than the Newey-West estimates. Results in Burnside and Eichenbaum (1996) suggest that using a relatively small number of moments and as much a priori information (e.g., model structure) as possible can improve coverage.

Canova, F. and Ferroni, F. 2008, Multiple Filtering Devices for the Estimation of Cyclical DSGE models, *mimeo.*

Canova, F. and Sala, L. 2009, Back to Square One: Identification Issues in DSGE Models, *Journal of Monetary Economics* **56**, 431–449.

Chernozhukov, V. and Hong, H. 2003, An MCMC approach to classical estimation, *Journal of Econometrics* **115**, 293 346.

Christiano, L. and den Haan, W. 1996, Small-Sample Properties of GMM for Business Cycle Analysis, *Journal 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., Eichenbaum, 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, *Journal of Economic Dynamics and Control* **25**, 1485–1525.

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

Cogley, T. and Nason, J. 1995b, Output dynamics in real business cycle models, *American Economic Review* **85**, 492511.

Coibion, O. and Gorodnichenko, Y. 2008, Strategic interaction among heterogeneous price-setters in an estimated DSGE model. NBER WP 14323.

Del Negro, M., Schorfheide, F., Smets, F. and Wouters, R. 2007, On the Fit of New Keynesian models, *Journal 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.

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

Fernandez-Villaverde, J. and Rubio-Ramirez 2006, Estimating Macroeconomic Models: A Likelihood Approach, *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, *Journal 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., Mikusheva, A. and Ng, S. 2009, A Simple root-T Consistent and Asymptotically Normal Estimator for the Largest Autoregressive Root, mimeo.

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

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

Horowitz, J. 1998, Bootstrap Methods for Covariance Structures, *Journal of Human Resources* **33**, 39–61.

Ireland, P. 1997, A Small Structural Quarterly Model for Monetary Policy Evaluaton, *Carnegie-Rochester Conference Series on Public Policy* **47**, 83108.

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

Ireland, P. 2004, 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, *Journal of Monetary Economics* **45**, 329–359.

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

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

Lubik, T. and Schorfheide, F. 2004, Testing for Indeterminacy: An Application to U.S. Monetary Policy, *American Economic Review* **94**, 190–217.

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.

Newey, W. and McFadden, D. 1994, Large Sample Estimators and Hypothesis Testing, *Handbook of Econometrics*, Vol. 4,Chapter 36, North Holland.

Ruge-Murcia, F. 2007, Methods to Estimate Dynamic Stochastic General Equilibrium Models, *Journal of Economic Dynamics and Control* **31**, 2599–2636.

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

Sims, C. 2002, Solving Linear Rational Expectations Models, *Computational Economics* pp. 1–20.

Singleton, K. J. 1988, Econometric Issues in the Analysis of Equilibrium Business Cycle Models, *Journal 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 Bayesian DSGE Approach, *American Economic Review* **97**, 586–606.

Wu, C. F. 1981, Asymptotic Theory of NonLinear Least Squares Estimation, *Annals of Statistics* **9:3**, 501–513.

9 Appendix

Consistency of the estimators follow from Wu (1981):

Lemma 1 *Let θ be the parameter of interest and let θ_0 denote the true value of θ. Suppose that for any $\delta > 0$*

$$\liminf_{T\to\infty} \inf_{\|\theta-\theta_0\|\geq\delta} (Q_T(\theta) - Q_T(\theta_0)) > 0 \text{ a.s.} \quad \text{or in probability.}$$

Then $\widehat{\theta}_T \xrightarrow{a.s.} \theta_0$ (or in probability) as $T \to \infty$.

From solving the baseline model, the endogenous variable y_t has solution

$$\begin{aligned} y_t &= v_{yk}k_{t-1} + v_{yz}z_t \\ k_t &= v_{kk}k_{t-1} + v_{kz}z_t \\ z_t &= \rho z_{t-1} + e_t. \end{aligned}$$

It is straightforward to show that y_t is an ARMA(2,1) with

$$y_t(1 - v_{kk}L)(1 - \rho L) = (1 - \theta_y L)e_t$$

where $\theta_y = v_{kk} - \alpha v_{kz}$. Note that v_{kk} does not depend on ρ, but v_{kz} does. A similar equation holds for consumption. Let the true parameters be denoted by a superscript 0 and to focus on the issue, we assume σ^2 and v_{kk} are known. Generically, write

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

where $b(\rho, v_{kk})$ is continuous in ρ and v_{kk}. Then the DGP is

$$y_t = \rho^0 y_{t-1} + u_t$$

where u_t is ARMA(1,1). Let $\widetilde{y}^0_t = (1 - \rho^0 L)y_t$ be y_t quasi-differenced at the true ρ and let $\widetilde{y}_t$ be y_t quasi-differenced at an arbitrary ρ. For any ρ that is assumed to be true, the analytical autocovariance at lag j for $\widetilde{y}_t$ is given by

$$\begin{aligned} \gamma_0(\rho) &= \sigma^2 \frac{1 + b(\rho, v_{kk})^2 + 2v_{kk}b(\rho, v_{kk})}{1 - v^2_{kk}} \\ \gamma_1(\rho) &= \sigma^2 \frac{(1 + v_{kk}b(\rho, v^0_{kk}))(v_{kk} + b(\rho, v_{kk}))}{1 - v^2_{kk}} \\ \gamma_j(\rho) &= v^{0,j-1}_{kk}\gamma_1(\rho), \quad j \geq 2 \end{aligned}$$

The derivatives of $\gamma_j(\rho)$ with respect to ρ exists and is well defined.

Let $\gamma_{Tj}(\rho)$ be the sample autocovariance of y quasi-transformed at an arbitrary ρ.

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

Without loss of generality, consider $j = 1$ and focus on the last term of the expression above. Now $y_{t-1} = \rho_0 y_{t-2} + u_{t-1}$. At $\rho^0 = 1$,

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}y_{t-2} - y^2_{t-1} = -\frac{1}{T}\sum_{t=1}^{T} y_{t-1}u_{t-1} = O_p(1).$$

It is easy to see that for any ρ in the $\frac{1}{\sqrt{T}}$ neighborhood of ρ_0,

$$\begin{aligned}\sqrt{T}(\gamma_{T1}(\rho)-\gamma_{T0}(\rho)) &= \left[T^{-1/2}\sum_{t=1}^{T}\widetilde{y}_t^0\widetilde{y}_{t-1}^0-(\widetilde{y}_t^0)^2\right] \\ &\quad -\sqrt{T}(\rho-\rho^0)\frac{1}{T}\sum_{t=1}^{T}\left[y_{t-1}\widetilde{y}_{t-1}^0+y_{t-2}\widetilde{y}_t^0-2y_{t-1}\widetilde{y}_t^0\right]+o_p(1)\end{aligned}$$

since $(\rho-\rho^0)^2T^{-1}\sum_{t=1}^{T}y_tu_t=O_p(T^{-1/2})$. The scaled sample orthogonality condition is

$$\begin{aligned}\sqrt{T}\bar{g}(\rho) &= \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\widetilde{y}_t^0\widetilde{y}_{t-1}^0-\gamma_1(\rho)\right)-\left((\widetilde{y}_t^0)^2-\gamma_0(\rho)\right)\right] \\ &\quad -\sqrt{T}(\rho-\rho^0)\frac{1}{T}\sum_{t=1}^{T}\left[y_{t-1}\widetilde{y}_{t-1}^0+y_{t-2}\widetilde{y}_t^0-2y_{t-1}\widetilde{y}_t^0\right]+o_p(1).\end{aligned}$$

Let $G_T(\rho)=\frac{1}{T}\sum_{t=1}^{T}\left[y_{t-1}\widetilde{y}_{t-1}^0+y_{t-2}\widetilde{y}_t^0-2y_{t-1}\widetilde{y}_t^0\right]$. The first order condition is $G_T'(\widehat{\rho})\bar{g}(\widehat{\rho})=0$. Evaluated at $\gamma_1(\rho_0)$ and $\gamma_0(\rho_0)$, the terms in the first square bracket obeys a central limit theorem. Let ξ_1 and ξ_0 be two independent normal random variables. Then

$$\sqrt{T}\bar{g}(\widehat{\rho})=\xi_1-\xi_0+G_T'(\rho^0)\sqrt{T}(\widehat{\rho}-\rho^0)+o_p(1)$$

Direct calculations yield

$$\sqrt{T}(\widehat{\rho}-\rho^0) = -(G_T'(\widehat{\rho})G_T(\rho^0))^{-1}G_T'(\rho^0)(\xi_1-\xi_0)+o_p(1).$$

Now $\xi_1-\xi_0$ has variance $2\sigma^4$ and

$$\begin{aligned}G_T(\rho) &= \frac{1}{T}\sum_{t=1}^{T}\left[y_{t-1}\widetilde{y}_{t-1}^0+y_{t-2}\widetilde{y}_t^0-2y_{t-1}\widetilde{y}_t^0\right] \\ &= \frac{1}{T}\sum_{t-1}^{T}(y_{t-2}+u_{t-1})u_{t-1}+y_{t-2}u_t-2y_{t-1}u_t\to\sigma^2\end{aligned}$$

since $T^{-1}\sum_{t=1}^{T}y_{t-2}u_{t-1}$, $T^{-1}\sum_{t=1}^{T}y_{t-1}u_t$ and $T^{-1}\sum_{t=1}^{T}y_{t-2}u_t$ all converge weakly to $\omega^2\int_0^1 W(r)dW(r)$, $W(r)$ is a standard Brownian motion, and ω^2 is the long run variance of u_t. Thus $\sqrt{T}(\widehat{\rho}-\rho^0)\xrightarrow{d}N(0,2)$. When other parameters are estimated, the asymptotic variance will be different but the developments are analogous. A local to unity framework can be used to show that $\widehat{\rho}$ is also $\sqrt{T}$ consistent when ρ^0 is in the local neighborhood of one. A more rigorous analysis is given in Gorodnichenko et al. (2009).

Table 1: Summary of Selected Work

Paper	Equations	Forcing variable	Model Filter	Data Filter	Estimator
Kydland and Prescott (1982)	system	ARMA(1,1)	LT	HP	calibration
Altug (1989)	system	I(1)	FD_1	FD_1	MLE
Christiano and Eichenbaum (1992)	system	I(1)	z_t	HP	GMM
Burnside et al. (1993)	system	AR(1)	LT	HP	GMM
Burnside and Eichenbaum (1996)	system	I(1)	z_t	HP	GMM
McGrattan et al. (1997)	system	VAR(2)	LT	LT,HP	MLE
Fuhrer (1997)	equation	not specified	not specified	HP,LT,QT	GMM
Clarida et al. (2000)	equation	AR(1)	not specified	LT,HP,CBO	GMM
Kim (2000)	system	AR(1)	LT	LT	MLE
Ireland (2001)	system	AR(1)	LT	LT	MLE
Smets and Wouters (2003)	system	AR(1)	LT	HP	Bayesian
Dib (2003)	system	AR(1)	LT	LT	MLE
Fuhrer and Rudebusch (2004)	equation	not specified	not specified	HP,CBO,QT	MLE,GMM
Lubik and Schorfheide (2004)	system	AR(1)	LT	HP,LT	Bayesian
Altig et al. (2004)	system	ARI(1,1)	FD_1	FD_1	GMM
Ireland (2004)	system	I(1)	FD_1	FD_1	MLE
Bouakez et al. (2005)	system	AR(1)	LT	LT	MLE
Christiano et al. (2005)	system	not specified	not specified	VAR	GMM
Del Negro et al. (2007)	system	ARI(1,1)	FD_1	FD_1	Bayesian
Faia (2007)	system	AR(1)	LT	HP	calibration
Smets and Wouters (2007)	system	AR(1)	FD	FD	Bayesian

Note: CBO denotes actual series minus the Congress Budget Office's measure of potential output. I(1) and ARI(1,1) denote forcing variables with stochastic trends. VAR, AR and ARMA denote trend stationary forcing variables. FD is first differencing, FD_1 is first differencing with the restriction that the forcing variable has a unit root (e.g., $\rho_z = 1$), LT is projection on linear time trend, QT is projection on quadratic time trend, HP is Hodrick-Prescott filter, z_t is detrending by the level of technology. The second column shows whether a paper estimates a system of equations ("system") or a single structural equation ("equation").

Table 2. Neoclassical Growth Model

ρ_z	data filter	QD	HD	FD	HP	LT	FD_1	HP	HP
	model filer	QD	HD	FD	HP	LT	FD_1	LT	z_t
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Estimate of α									
0.95	mean	0.318	0.333	0.367	0.350	0.480	0.400	0.675	0.990
	st.dev.	0.052	0.061	0.110	0.103	0.120	0.083	0.022	
0.99	mean	0.308	0.324	0.372	0.360	0.810	0.377	0.789	0.990
	st.dev.	0.053	0.066	0.115	0.120	0.201	0.109	0.024	
1.00	mean	0.304	0.312	0.349	0.351	0.905	0.357	0.817	0.990
	st.dev.	0.054	0.061	0.105	0.115	0.183	0.113	0.022	
Estimate of ρ_z									
0.95	mean	0.949	0.949	0.950	0.950	0.914	1.000	0.541	1.000
	st.dev.	0.006	0.014	0.017	0.015	0.042		0.049	
0.99	mean	0.989	0.990	0.991	0.991	0.864	1.000	0.485	1.000
	st.dev.	0.002	0.005	0.007	0.016	0.094		0.041	
1.00	mean	0.999	1.000	0.998	1.000	0.694	1.000	0.461	1.000
	st.dev.	0.001	0.003	0.005	0.011	0.123		0.039	
Estimate of σ_z									
0.95	mean	0.981	1.021	1.157	1.076	1.135	1.334	1.949	0.046
	st.dev.	0.123	0.187	0.441	0.291	0.283	0.285	0.167	0.006
0.99	mean	0.962	1.001	1.154	1.107	4.348	1.185	2.912	0.042
	st.dev.	0.111	0.170	0.367	0.303	2.169	0.347	0.397	0.006
1.00	mean	0.955	0.974	1.073	1.087	19.803	1.107	3.289	0.041
	st.dev.	0.108	0.145	0.295	0.513	10.681	0.341	0.478	0.005

Note: The number of simulations is 2,000. Sample size is T=200. LT is linear detrending, HP is Hodrick-Prescott filter, FD is first differencing, FD_1 is first differencing with the restriction that $\rho_z = 1$, QD is quasi differencing, HD is hybrid differencing, z_t is detrending by the level of technology.

Table 3. Augmented Versions of the Neoclassical Growth Model

ρ_z	data filter / model filer	QD / QD (1)	HD / HD (2)	FD / FD (3)	HP / HP (4)	LT / LT (5)	FD_1 / FD_1 (6)	HP / LT (7)	HP / z_t (8)
	Panel A: serially correlated growth rate in technology								
	Estimate of $\kappa = 0$								
0.95	mean	-0.010	-0.001	0.001	-0.019	-0.180	-0.100	-0.224	-0.369
	st.dev.	0.063	0.058	0.050	0.160	0.165	0.035	0.038	0.106
0.99	mean	-0.014	-0.003	0.002	-0.016	-0.498	-0.021	-0.255	-0.429
	st.dev.	0.044	0.046	0.041	0.155	0.088	0.030	0.038	0.070
1.00	mean	-0.014	-0.003	0.003	-0.020	-0.600	-0.002	-0.256	-0.446
	st.dev.	0.038	0.039	0.035	0.161	0.029	0.028	0.038	0.058
	Panel B: habit formation in consumption								
	Estimate of $\phi = 0$								
0.95	mean	0.020	0.008	0.006	-0.014	-0.410	-0.086	0.193	0.679
	st.dev.	0.075	0.073	0.071	0.255	0.339	0.066	0.382	0.074
0.99	mean	0.019	0.009	0.008	0.023	-0.647	-0.018	0.495	0.637
	st.dev.	0.070	0.074	0.069	0.168	0.241	0.083	0.400	0.070
1.00	mean	0.018	0.023	0.011	0.025	-0.702	0.011	0.603	0.622
	st.dev.	0.067	0.087	0.076	0.156	0.174	0.095	0.373	0.068
	Panel C: preference shocks q_t								
	Estimate of $\alpha = 0.33$								
	$\sigma_q = 0.5$								
0.95	mean	0.344	0.335	0.361	0.329	0.465	0.299	0.591	0.337
	st.dev.	0.040	0.025	0.080	0.075	0.126	0.043	0.037	0.137
1.00	mean	0.353	0.342	0.352	0.335	0.508	0.352	0.665	0.485
	st.dev.	0.052	0.034	0.067	0.072	0.357	0.063	0.050	0.265
	$\sigma_q = 1.0$								
0.95	mean	0.339	0.341	0.349	0.331	0.431	0.316	0.504	0.344
	st.dev.	0.023	0.030	0.049	0.060	0.103	0.022	0.031	0.022
1.00	mean	0.347	0.347	0.357	0.340	0.611	0.341	0.529	0.364
	st.dev.	0.028	0.034	0.050	0.055	0.257	0.025	0.036	0.023
	$\sigma_q = 1.5$								
0.95	mean	0.338	0.343	0.349	0.333	0.399	0.326	0.469	0.378
	st.dev.	0.021	0.030	0.042	0.053	0.078	0.017	0.020	0.024
1.00	mean	0.341	0.346	0.353	0.338	0.515	0.338	0.477	0.391
	st.dev.	0.020	0.029	0.038	0.049	0.203	0.017	0.021	0.023

Note: Panels A and B: other parameters fixed at $\alpha = 0.33$ and $\sigma_z = 1$. Panel C: five parameters are estimated $(\alpha, \rho_z, \rho_q, \sigma_z, \sigma_q)$. The number of simulations is 2,000. Sample size is T=200. LT is linear detrending, HP is Hodrick-Prescott filter, FD is first differencing, FD_1 is first differencing with the restriction that $\rho_z = 1$, QD is quasi differencing, HD is hybrid differencing, z_t is detrending by the level of technology.

Table 4. Smets and Wouters (2007) model.

ρ_z	data filter model filer	QD QD (1)	HD HD (2)	FD FD (3)	HP HP (4)	LT LT (5)	FD_1 FD_1 (6)	HP LT (7)	HP z_t (8)
	Estimate of persistence in technology shocks ρ_z								
0.95	mean	0.965	0.967	0.962	0.945	0.864	1.000	-0.100	1.000
	st.dev.	0.038	0.037	0.044	0.137	0.142		0.157	
0.99	mean	0.986	0.984	0.986	0.967	0.836	1.000	-0.114	1.000
	st.dev.	0.027	0.027	0.028	0.123	0.227		0.090	
1.00	mean	0.990	0.989	0.993	0.971	0.744	1.000	-0.123	1.000
	st.dev.	0.027	0.026	0.025	0.123	0.305		0.075	
	Estimate of investment adjustment cost $\phi = 5.48$								
0.95	mean	5.057	5.381	5.227	5.066	3.932	4.700	4.447	9.818
	st.dev.	2.236	2.548	2.306	3.354	1.917	2.487	0.265	0.609
0.99	mean	5.432	5.563	5.373	5.095	5.595	5.236	4.366	9.662
	st.dev.	2.321	2.463	2.404	3.012	2.647	2.794	0.257	0.588
1.00	mean	5.863	6.253	6.014	5.617	6.173	6.049	4.377	9.541
	st.dev.	2.375	2.775	2.781	3.279	2.983	3.046	0.230	0.548
	Estimate of habit formation $\lambda = 0.71$								
0.95	mean	0.725	0.730	0.749	0.753	0.730	0.864	3.932	0.673
	st.dev.	0.057	0.063	0.062	0.049	0.063	0.142	1.917	0.134
0.99	mean	0.699	0.718	0.719	0.718	0.543	0.744	0.908	0.941
	st.dev.	0.056	0.053	0.062	0.134	0.177	0.053	0.033	0.006
1.00	mean	0.686	0.711	0.716	0.709	0.470	0.731	0.912	0.940
	st.dev.	0.056	0.055	0.064	0.145	0.261	0.057	0.028	0.005
	Estimate of wage adjustment probability $\xi_w = 0.73$								
0.95	mean	0.704	0.730	0.734	0.686	0.657	0.759	0.484	0.220
	st.dev.	0.073	0.063	0.075	0.117	0.105	0.077	0.085	0.019
0.99	mean	0.686	0.704	0.709	0.659	0.530	0.718	0.458	0.213
	st.dev.	0.081	0.065	0.079	0.125	0.214	0.084	0.078	0.016
1.00	mean	0.673	0.697	0.700	0.641	0.457	0.700	0.444	0.210
	st.dev.	0.092	0.068	0.083	0.138	0.262	0.091	0.072	0.015

Note: The number of simulations is 2,000. Sample size is T=150. LT is linear detrending, HP is Hodrick-Prescott filter, FD is first differencing, FD_1 is first differencing with the restriction that $\rho_z = 1$, QD is quasi differencing, HD is hybrid differencing, z_t is detrending by the level of technology.

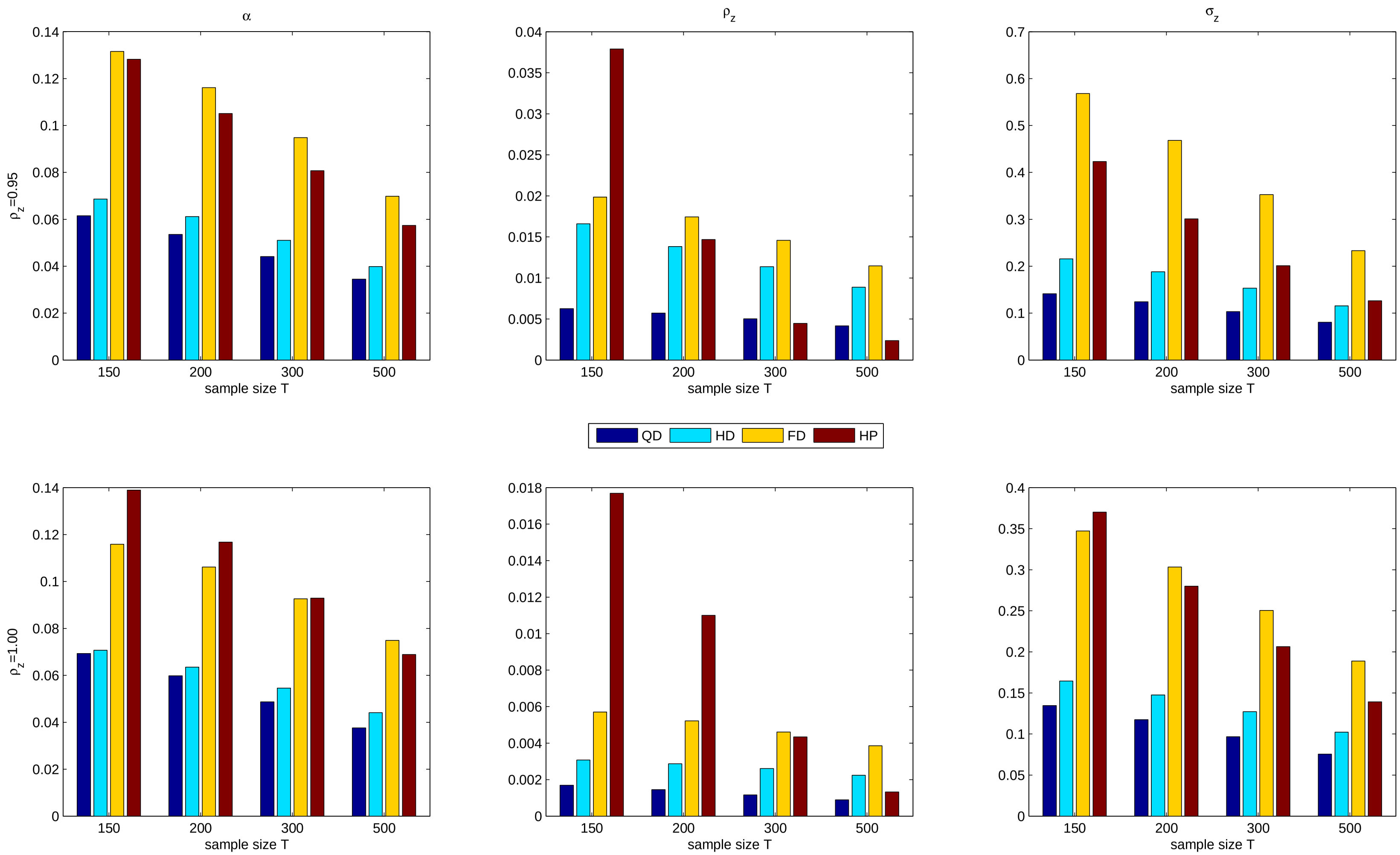

Figure 1: Root mean squared errors.

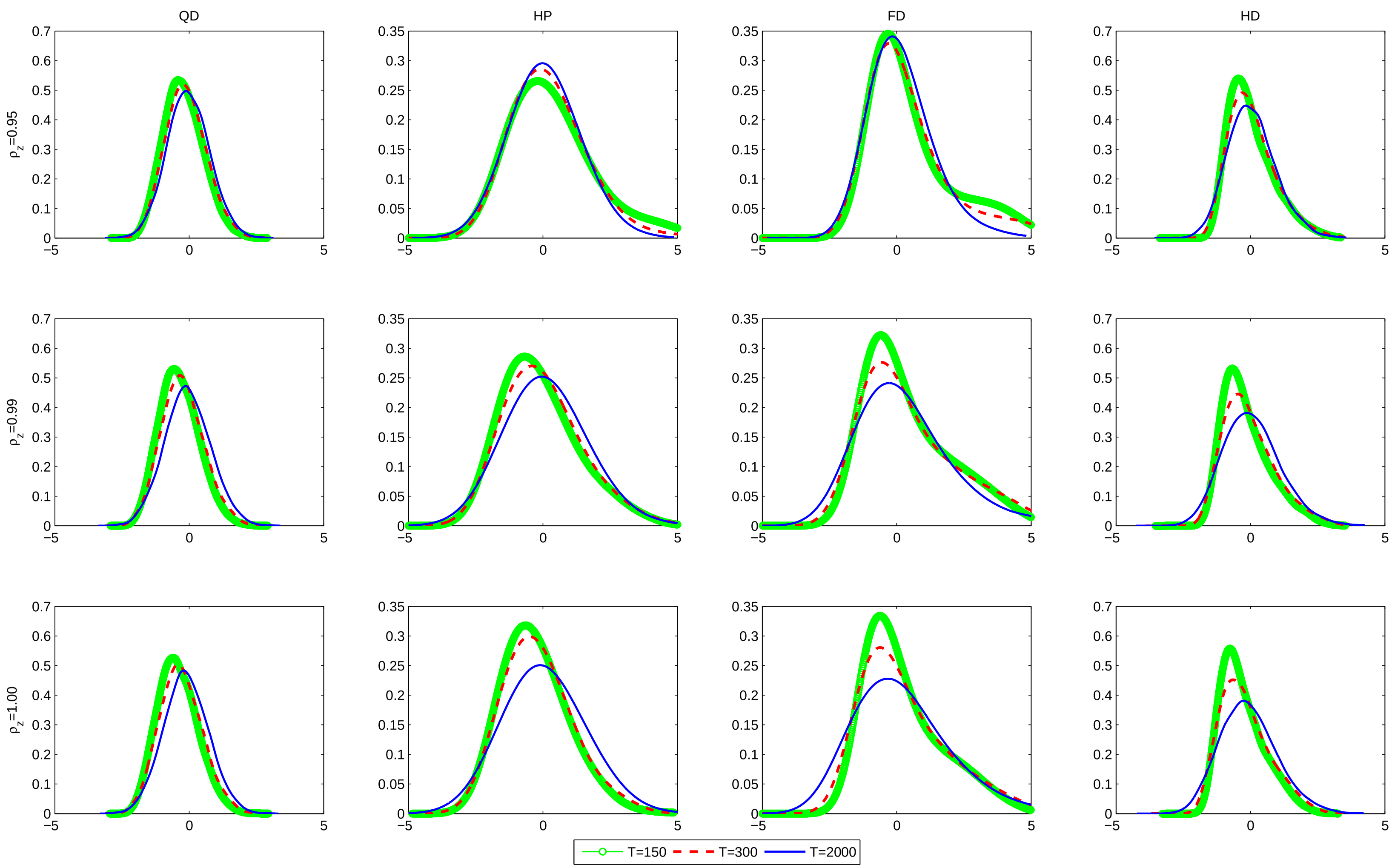

Figure 2: Kernel density of simulated $\sqrt{T}(\widehat{\alpha} - \alpha)$.

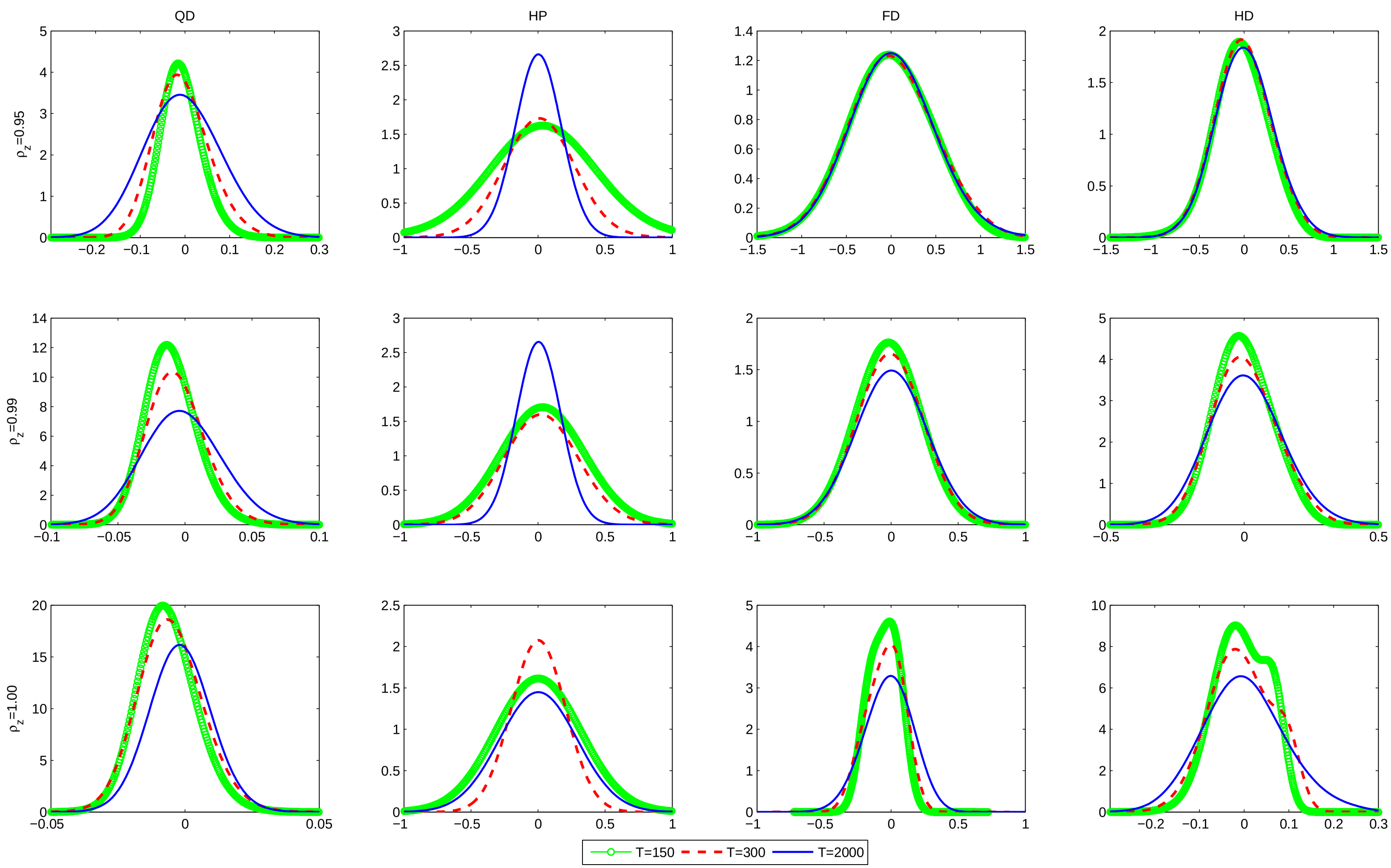

Figure 3: Kernel density of simulated $\sqrt{T}(\widehat{\rho}_z - \rho_z)$.

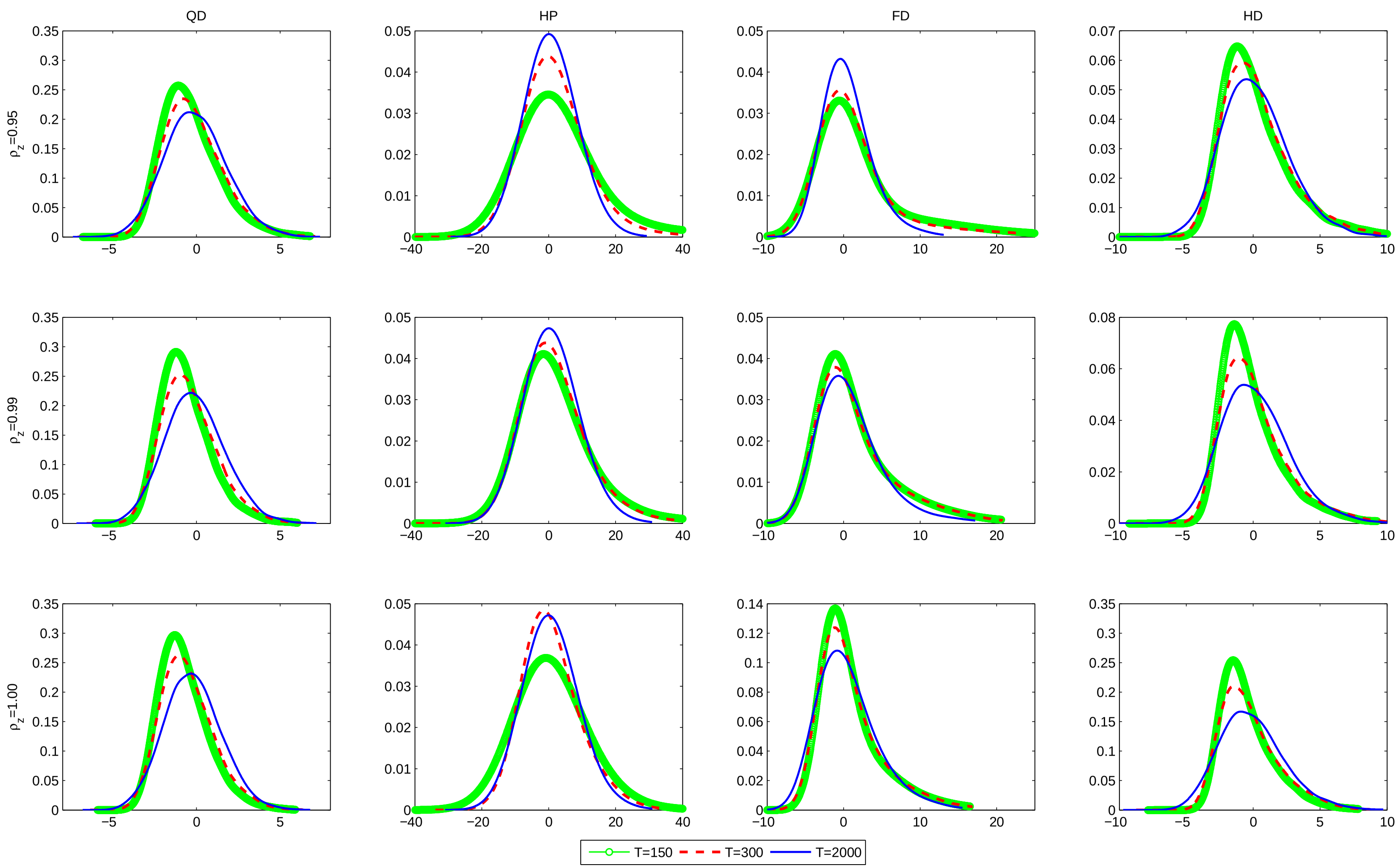

Figure 4: Kernel density of simulated $\sqrt{T}(\widehat{\sigma}_z - \sigma_z)$.

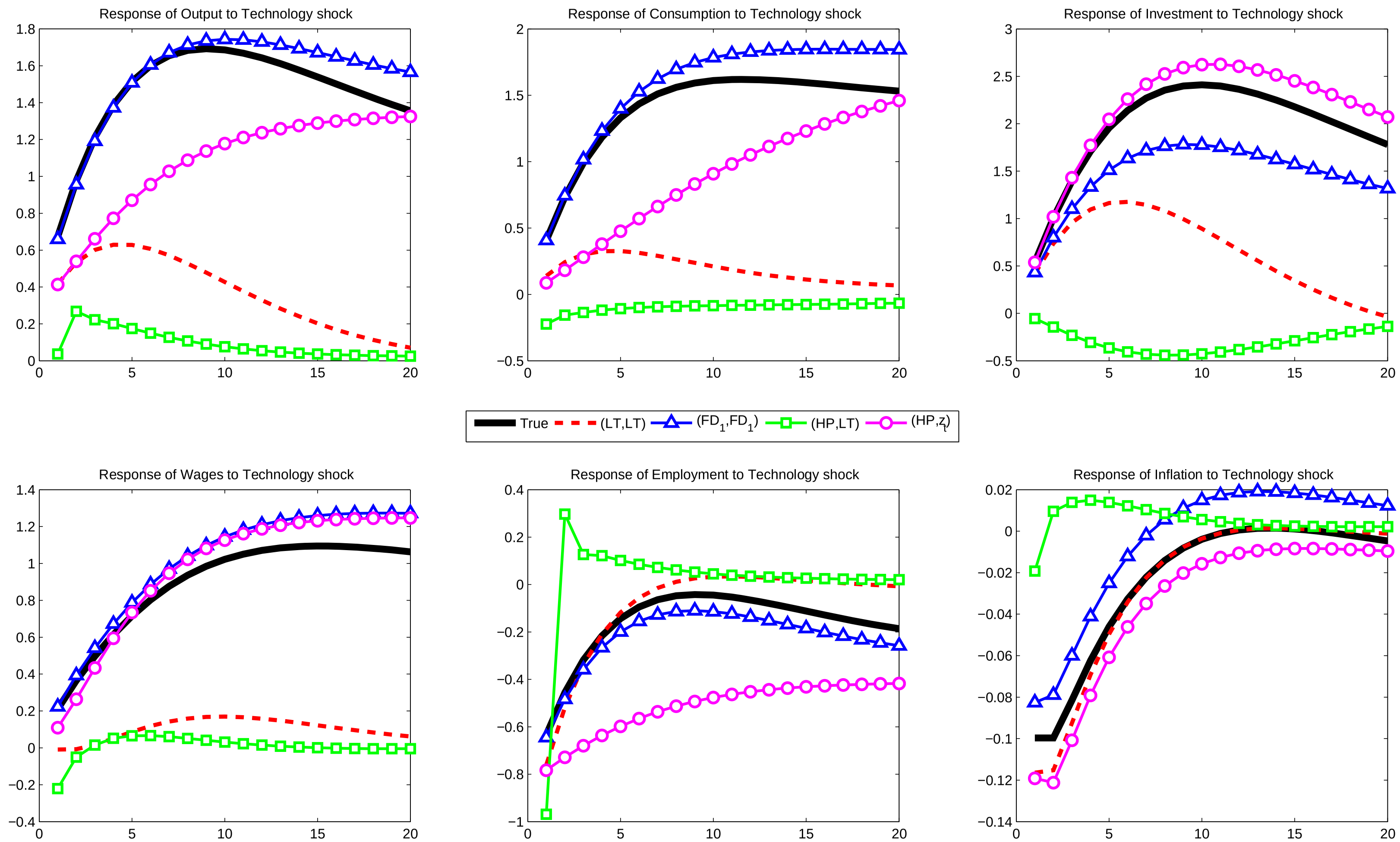

Figure 5: Impulse responses functions in the estimated Smets and Wouters (2007) model. Persistence of technology shock is $\rho_z = 0.99$.

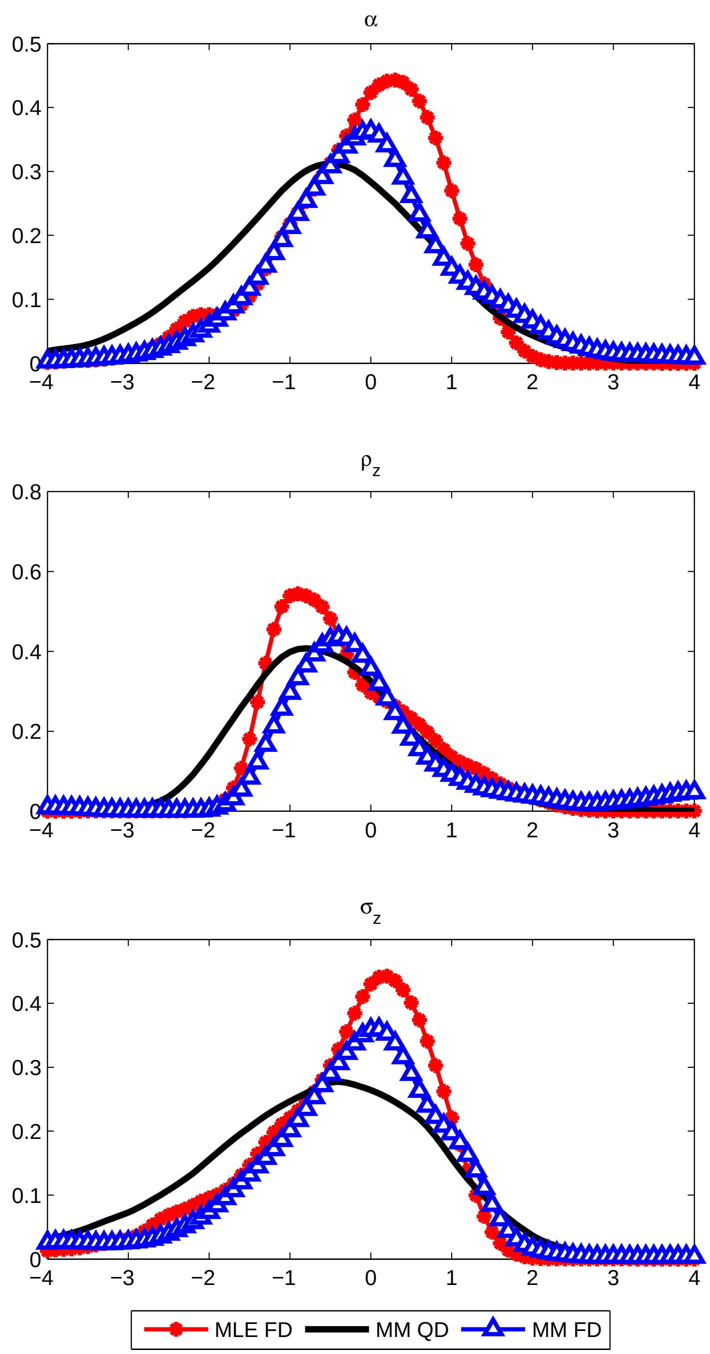

Figure 6: Density of simulated t-statistic for MLE and MM estimators.