MELBOURNE INSTITUTE
Applied Economic & Social Research

Working Paper Series

Too Many Shocks Spoil the Interpretation

Adrian Pagan
Tim Robinson

Working Paper No. 2/20
April 2020

THE UNIVERSITY OF MELBOURNE
MELBOURNE INSTITUTE

Too Many Shocks Spoil the Interpretation*

Adrian Pagan
School of Economics, University of Sydney
CAMA, Australian National University

Tim Robinson
Melbourne Institute: Applied Economic & Social Research,
The University of Melbourne

Melbourne Institute Working Paper No. 2/20
April 2020

* Some of the material appeared in an earlier working paper, Pagan and Robinson (2019). Our thanks to Andrew Harvey, George Kapetanios, Mariano Kulish, Dan Rees and Guido Turnip for useful comments. This work was supported by ARC Grant DP160102654.

Melbourne Institute: Applied Economic & Social Research
The University of Melbourne
Victoria 3010 Australia
Telephone **+61 3 8344 2100**
Fax **+61 3 8344 2111**
Email **melb-inst@unimelb.edu.au**
Website **melbourneinstitute.unimelb.edu.au**

Abstract

We show that when a model has more shocks than observed variables the estimated filtered and smoothed shocks will be correlated. This is despite no correlation being present in the data generating process. Additionally the estimated shock innovations may be autocorrelated. These correlations limit the relevance of impulse responses, which assume uncorrelated shocks, for interpreting the data. Excess shocks occur frequently, e.g. in Unobserved-Component (UC) models, filters, including Hodrick-Prescott (1997), and some Dynamic Stochastic General Equilibrium (DSGE) models. Using several UC models and an estimated DSGE model, Ireland (2011), we demonstrate that sizable correlations among the estimated shocks can result.

JEL classification: E37; C51; C52.

Keywords: Partial Information; Structural Shocks; Kalman Filter; Measurement Error; DSGE.

1 Introduction

The isolation of shocks, and the analysis of economic outcomes using them, is a major focus of macroeconomics. This approach is also prevalent in many other parts of economics. Dynamic Stochastic General Equilibrium (DSGE) and Structural Vector AutoRegression (SVAR) models are two of the workhorse approaches of modern macroeconomics and both focus on shocks. A crucial assumption in such analyses is a lack of contemporaneous correlation between shocks. If this does not hold one cannot say which shock is providing the explanation. Suppose, for example, that the monetary and fiscal policy shocks were correlated. In this circumstance we cannot isolate the impact of a monetary policy shock; instead we have to vary a *package* of shocks, since one shock cannot be changed without moving the other. These "packages" might be described as "mongrel shocks" - Canova and Ferroni (2019) - as they have no clear interpretation. The contribution of this paper is to demonstrate instances when these mongrel shocks *must* occur.

Estimated DSGE and SVAR models initially had the number of model shocks equaling the number of observed variables; a noteworthy example being Smets and Wouters (2007). It was recognized that having too few shocks was an issue, as that would imply a singularity in the covariance matrix of observable variables. Such a restriction enabled estimation via maximum likelihood. The likelihood was constructed using the one-step prediction errors of the observed variables, namely $\eta_t = y_t - \mathbb{E}_{t-1}(y_t)$, where $\mathbb{E}_{t-1}$ is the expectation of a vector of observables y_t conditioned on their past history. These predictions errors could be readily obtained from the Kalman filter once the model was placed in a state-space form.

More recently, there have been an increasing number of DSGE models where the number of model shocks exceeds the number of observable variables, a situation we will refer to as *excess shocks*. The model shocks are often taken to be autoregressive processes. Consequently it is the innovations to these shocks that need to be uncorrelated for there to be a clear interpretation. They must also have no serial correlation; otherwise they are not innovations. Moreover, in addition to model shocks that have an economic interpretation other shocks are often included so as to improve the model's fit to the data. Three examples show this:

1. Measurement errors. The model is $y_t = y_t^* + \varepsilon_{1t}$ where y_t^* is a latent variable driven by a shock ε_{2t} which might be called (say) technology. Here ε_{1t} may be a measurement error and there are two shocks ε_{jt}, $j = 1, 2$, for a single observable y_t.

2. Indeterminancy. Here there are a certain number of model shocks when the solution is determinate and these could equal the number of observable variables, but when there is indeterminancy sunspot shocks could enter the solution, and so the total number of shocks could exceed the observables.

3. News shocks. In these cases one generally ends up with more shocks than observables, e.g. in Christiano et al. (2014) there are 12 observed variables but 20 model shocks, the difference being due to the presence of eight news shocks. The news shocks are correlated with each other but they have a non-singular covariance matrix, i.e. do not have perfect correlation.

Beyond DSGE models, there are many other examples:

1. Markov Switching. Here the model is $y_t = a + bz_t + \varepsilon_{1t}$ and the z_t is a latent Markov process driven by a binary shock. Often this is applied to a single observed series y_t while in this case there are two shocks.

2. Unobserved-Components (UC) models. These decompose a single series into two components by using a model for the components. Often these components are called trend and cycle, or permanent and transitory. A classic example is the derivation of the Hodrick-Prescott (HP) filter (Hodrick and Prescott 1997). Placed in state-space form and estimated the shocks produced are the estimated trend and cyclical components. Again there are more shocks than observed variables.

3. Time-varying parameters. The simplest of these would be $y_t = \beta_t + \varepsilon_{1t}$ where β_t is stochastically varying and driven by an extra shock ε_{2t}.

4. Stochastic-volatility models. For example, let $y_t = \varepsilon_{1t}\sigma_t$ with $\Delta \log \sigma_t^2 = \varepsilon_{2t}$. Then $\Delta(\log y_t^2) = \Delta \log \varepsilon_{1t}^2 + \varepsilon_{2t}$ and there are two shocks but one observable.

Now in the estimation of these models the assumptions made about the innovations into the shocks are held to be correct. If the model is incorrect then the assumptions above may be false for the measured shock innovations emerging after estimation. To avoid such specification issues in this paper we will frequently assume that the assumptions made about the model shocks are actually correct, i.e. the Data Generating Process (DGP) used to describe the shocks is the correct one. We might then expect that in large samples, and no excess shocks, the shock innovations estimated from the data would have the properties of no contemporaneous and serial correlation, provided the model is identified. The concern of this paper is whether this remains true when there are more shocks than observable variables used in estimation.

We show that when there are more shocks than observed variables, the measured shock innovations - those obtained using the data - do *not* have the same properties as the model shock innovations. In the event that the shocks are constructed using contemporaneous data, i.e. are *filtered*, then their innovations have a singular multivariate spectrum - unlike the spectrum of the innovations in the DGP - and this shows up as bivariate *contemporaneous correlations.* Additionally, the innovations may be serially correlated. These results hold regardless of whether the shocks are filtered or smoothed, although the correlations can differ quantitatively. Essentially, when excess shocks exist mongrel shocks *will* occur.

It should be said from the outset that this is *not* an issue with the estimation methodology or identification. It may be that the model is unidentified, even when there are no excess shocks. Nor is it a consequence of sample size. It is an issue of *partial information.* Indeed, to avoid complexities such as weak identification we will frequently assume that the innovations to the model shocks are uncorrelated and all parameters of the model are known. The issues that arise are those of shock *separation.* If all one wants to do is to estimate a model and test various restrictions then this can be done with excess shocks (ignoring any identification problems). However, excess shocks end up causing the measured shock innovations to have properties that the model shock innovations do *not* have. This is particularly problematic for models whose properties are interpreted using these shocks by examining items such as impulse responses. An example is the use of impulse responses to understand the dynamics of DSGE models.

In Section 2 we demonstrate that it is simple to show that the excess shocks will imply a singular covariance matrix between the estimated shocks when the underlying dynamic structure is a VAR and all variables are observed. However, many of the models mentioned above that produce an excess of shocks have unobserved variables and therefore need to be formulated in a state-space form or its equivalent. Consequently, we show in Section 3 that the measured innovations to shocks found with either the Kalman filter or smoother also have different properties to those of the model shock innovations. Specifically, the filtered innovations have a singular spectral density and the smoothed shock innovations are contemporaneously and, almost certainly, serially correlated.

The presence of correlation in the measured shocks is true not only of models estimated using the Kalman filter, but also of other filters that produce $\mathbb{E}_{t-1}(y_t)$. It is important to recognize this as there appears to be a perception that filters can separate the shocks even when an excess exists; certainly graphs are often presented of the contribution of individual shocks to data and techniques such as impulse responses or variance decompositions are used. These require contemporaneously and serially uncorrelated innovations, properties which the data-based shocks do not have, unlike the model shocks of the DGP.

The fundamental problem is that while the Kalman filter, for example, enables the computation of an estimate of η_t, this is a combination of the data and *all* the shocks (and their lags). Consequently, knowledge of η_t alone cannot be used to separate the shocks when there are more shocks than observables. It may be possible to separate *some* of the shocks, but not all of them. Unfortunately this is case-dependent and always needs to be checked.

Section 4 presents several examples, initially allowing the model variables to be I(0), and subsequently having a mixture of I(1) and I(0) variables. We find that many unobserved component models where shock innovations are assumed to be uncorrelated (and are that in the DGP) result in measured innovations that are correlated both contemporaneously and serially. For example, this is the case for both the one-sided and two-sided HP filter. It has often been noted that the UC model that can deliver the HP filter assumes contemporaneously and serially uncorrelated innovations into the permanent and transitory components, and yet the estimated transitory component shock has a large amount of serial correlation. We show that even when the DGP has uncorrelated innovations excess shocks will result in the estimated innovations having serial correlation, and that the filtered trend and cycle innovations are perfectly correlated.

We present an empirical example of a small-scale DSGE model from the literature, Ireland (2011), to demonstrate the magnitudes of the correlations among the empirical innovations which may result from having excess shocks. These correlations can be sizable, making interpretation of the shocks difficult.

Finally, in Section 5 we address several potential issues that may be raised about excess shocks. In particular, we show that excess shocks are often a consequence of introducing measurement error; that the associated correlations will occur regardless of whether maximum likelihood or Bayesian methods are used; and this holds regardless of whether there is a small or large sample of data. Section 6 concludes.

2 Partial Information and Structural Shock Estimation for SVAR-Type Structures

Excess shocks arise due to partial information, i.e. there are not enough observables to separate the shocks. Consider a $n \times 1$ vector of observable variables y_t that relate to a $m \times 1$ vector of shocks ε_t as $y_t = A\varepsilon_t$. Here ε_t are the model shocks. We assume that an excess of shocks exists, i.e. $m > n$, A is known and ε_t is an innovation with the properties $\varepsilon_t \sim N(0, I_m)$, and $\mathbb{E}(\varepsilon_t \varepsilon'_{t-j}) = 0,\ j > 0$.

The measured shock innovations need to be found from the data. Generally we want to compute these with some information set F as $\mathbb{E}(\varepsilon_t|F)$. If we use information up to t then we would have *filtered* estimates $E_t\varepsilon_t = \mathbb{E}(\varepsilon_t|y_1...y_t)$, while *smoothed* estimates are $E_T\varepsilon_t = \mathbb{E}(\varepsilon_t|y_1...y_T)$, where the latter use all the data. In this case, with only observed variables, $\mathbb{E}(\varepsilon_t|F)$ is the same for $F = t$ or T, and so we will designate the measured shock innovations as $E_t\varepsilon_t$, while ε_t will be the model shock innovations.

There is no unique solution for $E_t\varepsilon_t$ in this instance. One solution is to use the generalized inverse to produce $E_t\varepsilon_t = A^+y_t$, where A^+ is a g-inverse (that is, A^+ is $m \times n$ and satisfies $AA^+A = A$) whose rank must be the smaller of m or n.[1] In practice of course we would have to estimate A, and then $E_t\varepsilon_t = \hat{A}^+y_t$, where the ˆ denotes it is an estimate, but we have assumed that A is known. It follows that the $var(E_t\varepsilon_t)$ must be singular since it is $A^+\Omega A^+$, $\Omega = cov(y_t)$ only has rank n, and the rank of the product of two matrices is the minimum of the rank of each, so n is the maximum possible rank and it is less than m. Such a circumstance is probably what led Ravenna (2007), in the context of understanding the relationship between DSGE and SVAR models, to state "..it will not be possible to map y_t into a higher-dimension vector of orthogonal shocks" (p. 2051).

This situation produces mongrel shocks. The $E_t\varepsilon_t$ must be correlated, so as to reduce the rank of the covariance matrix of it down to that of the covariance matrix of y_t. Accordingly, variance decompositions which assume that the shocks are uncorrelated have no direct connection with the data as the *empirical shocks are correlated* due to a singularity in their covariance matrix. In a similar vein, in practice researchers often *assume* that $E_t\varepsilon_t$ have the same properties as ε_t when presenting impulse responses, and so vary an element of $E_t\varepsilon_t$ separately from the other shocks to examine its impact on the system. However, this is a *hypothetical experiment* relevant to the DGP model shocks, not those coming from the data, as one cannot vary elements of $E_t\varepsilon_t$ without changing others.

[1] The Moore-Penrose g-inverse additionally requires that A^+ satisfies $A^+AA^+ = A^+$, $(AA^+)^H = AA^+$, and $(A^+A)^H = A^+A$, where H denotes the conjugate transpose; see Campbell and Meyer (1991). When $m > n$ the Moore-Penrose g-inverse minimizes the Euclidean norm of the error $y_t - A\varepsilon_t$, but other norms would be possible.

Note that the same result applies if y_t instead follows a VAR(1):

$$y_t = B_1 y_{t-1} + A\varepsilon_t,$$

where B_1 are coefficients. In this case the smoothed and filtered shocks are again the same so that $E_t\varepsilon_t = A^+\eta_t$; recall η_t are the one-step prediction errors. Again the covariance matrix of $E_t\varepsilon_t$ is singular even if that of η_t is not. Allowing some of the ε_t to be correlated during estimation does not solve the problem; in order to match the singularity of the covariance matrix of $E_t\varepsilon_t$ it would be necessary for the covariance matrix of ε_t would also need to be singular, and this would mean that one of the model shocks should be deleted. The only way to avoid the singularity is to increase the dimension of y_t without changing the number of shocks.

3 Partial Information and Shock Estimation for State-Space Forms

Many estimated models do not have the structure of a finite-order VAR. Generally this occurs as a result of unobservable variables, which are effectively replaced by combinations of the lags of all the observables, and this is done by using a State-Space Form (SSF). We now consider this case, using the following SSF:

$$z_t = D_1\psi_t + D_2\psi_{t-1} + R\varepsilon_t \tag{1}$$

$$\psi_t = M\psi_{t-1} + C\varepsilon_t. \tag{2}$$

Here Equations (1) and (2) are the observation and state equations respectively, z_t are the $n \times 1$ observed variables, ψ_t the $p \times 1$ core model variables, and ε_t the $m \times 1$ vector of shock innovations which are assumed to be $N(0, I_m)$. By "core" model variables we mean those that cannot be substituted out. DSGE models, for example, often have variables that can be substituted out using identities included in the model. It must be that $p \geq n$ and with excess shocks $m > n$. Also, to reiterate, in DSGE models the shocks frequently are autocorrelated; we are focussing on their innovations. In most cases $p > m$. Nimark (2015) shows that the filtered estimate of ψ_t, $E_t\psi_t$, given the data to time t, and the system in Equations (1) and (2), evolves as

$$
\begin{aligned}
E_t\psi_t &= \Phi_t E_{t-1}\psi_{t-1} + K_t z_t && (3)\\
K_t &= [MP_{t-1|t-1}\Psi' + CC'D_1' + CR'][\Psi P_{t-1|t-1}\Psi' + \Lambda\Lambda']^{-1} && (4)\\
P_{t|t} &= P_{t|t-1} - K_t[\Psi P_{t-1|t-1}\Psi' + \Lambda\Lambda']K_t' && (5)\\
P_{t+1|t} &= MP_{t|t}M' + CC' && (6)\\
\Psi &= D_1 M + D_2, \Lambda = D_1 C + R, \Phi_t = M - K_t\Psi, && (7)
\end{aligned}
$$

where K_t is the gain of the Kalman filter, $P_{t|t}$ the filtered variance of $E_t\psi_t - \psi_t$ and $P_{t+1|t}$ provides the one-step-ahead predictor of the latter.[2] Again we see from Equation (3) that the estimated states are combinations of z_t, and so the matrix K_t does not have rank equal to $\dim(\psi_t)$.

When all model variables are observables $D_1 = I, D_2 = 0$ and $R = 0$, meaning that Equations (1) and (2) become

$$z_t = Mz_{t-1} + CE_t\varepsilon_t,$$

and $var(E_t\varepsilon_t)$ is singular if $\dim(E_t\varepsilon_t) > \dim(z_t)$. Recall $E_t\varepsilon_t = \mathbb{E}(\varepsilon_t|y_1...y_t)$ and denotes the filtered innovations.

Moving to the general case where there are unobservable variables we will assume, for simplicity, that either there is an infinite sample of data available or the parameters known,

[2]The variant of the Kalman filter used above, developed by Nimark (2015), allows for a lag of the state in the measurement equations. This is useful for the models which include permanent shocks, such as we study in Sub-section 4.2. Alternatively, one could expand the state to include a lag and use the standard Kalman filter recursions.

so that we can use the steady-state Kalman filter. Then Equation (3) tells us that

$$E_t\psi_t = \sum_{j=0} \Phi^j K z_{t-j}.$$

Denoting $\mathbb{E}_t(\psi_t)$ as $\psi_{t|t}$, from Equation (2) we obtain an expression for the filtered estimates of the shock innovations, namely

$$\psi_{t|t} = M\psi_{t-1|t} + CE_t\varepsilon_t.$$

Following Kurz (2018, Section 4)

$$\begin{aligned}
\psi_{t-1|t} &= \psi_{t-1|t-1} + P_{t-1|t-1}\Psi' F_t^{-1}(z_t - \Psi\psi_{t-1|t-1}) \\
F_t &= [\Psi P_{t-1|t-1}\Psi' + \Lambda\Lambda'].
\end{aligned}$$

Hence

$$\psi_{t|t} = M(\psi_{t-1|t-1} + P_{t-1|t-1}\Psi' F_t^{-1}(z_t - \Psi\psi_{t-1|t-1})) + CE_t\varepsilon_t.$$

In steady state

$$\begin{aligned}
\psi_{t|t} &= M\psi_{t-1|t-1} + MP\Psi' F^{-1}(z_t - \Psi\psi_{t-1|t-1})) + CE_t\varepsilon_t \\
&= M(I - P\Psi' F^{-1}\Psi)\psi_{t-1|t-1} + MP\Psi' F^{-1} z_t + CE_t\varepsilon_t \\
&= A\psi_{t-1|t-1} + Bz_t + CE_t\varepsilon_t,
\end{aligned}$$

with A and B implicitly defined. As $E_t\psi_t = \psi_{t|t}$, using Equation (3) this gives an expression for the filtered innovations $E_t\varepsilon_t$ as the solution to

$$\sum_{j=0} \Phi^j K z_{t-j} = A \sum_{j=1} \Phi^{j-1} K z_{t-j} + Bz_t + CE_t\varepsilon_t,$$

that is,

$$
\begin{aligned}
E_t\varepsilon_t &= \sum_{j=0} \Gamma_j z_{t-j} &(8)\\
\Gamma_0 &= C^+(K+B), \Gamma_1 = C^+(\Phi - A)K, \Gamma_2 = C^+(\Phi - A)\Phi K...
\end{aligned}
$$

The spectral density of $E_t\varepsilon_t$ is then $|\Gamma(\lambda)^2|f_{yy}(\lambda)$ and has rank equal to $f_{yy}(\lambda)$. When there are more shocks than variables it is rank deficient. Equation (8) also suggests that the filtered innovations may have serial correlation.

The central issue can be seen from looking at the one-step prediction error for z_t, η_t. From Kurz (2018, Equation 4.7), this is

$$
\begin{aligned}
\eta_t &= z_t - \mathbb{E}_{t-1}(z_t)\\
&= \Psi\phi_{t-1} + (D_1 C + R)\varepsilon_t,
\end{aligned}
$$

where $\phi_t = \psi_t - E_t\psi_t$. His Equation (4.8) gives an expression for ϕ_t

$$
\begin{aligned}
\phi_t &= (M - K_t\Psi)\phi_{t-1} + (C - K_t(D_1 C + R))\varepsilon_t &(9)\\
&= J_{1t}\phi_{t-1} + J_{2t}\varepsilon_t. &(10)
\end{aligned}
$$

Hence η_t is a linear combination of all the shocks ε_{t-j}. If one used a steady-state Kalman filter then J_{1t} and J_{2t} will be constant. Now the log likelihood depends directly on η_t so that, once we know $\eta_1, ..., \eta_T$, we know the likelihood. Because η_t depends in a linear form on $\{\varepsilon_k\}_{k=1}^t$, when there is an excess of model shocks, i.e. more ε_t than observables y_t, we would need to recover $E_t\varepsilon_t$ with a g-inverse, as in the VAR case. This means that the covariance matrix of $E_t\varepsilon_t$ is singular and it leads to correlation between the empirical shocks.

Kurz (2018, equation 4.11) shows that the smoothed states are obtained from the recursion

$$
\begin{aligned}
E_T\psi_t &= \psi_{t|T} = \psi_{t|t} + P_{t|t}\tau_t, &(11)\\
\tau_t &= G_1\eta_{t+1} + G_2\tau_{t+1}, &(12)
\end{aligned}
$$

where G_1 and G_2 are functions of the SSF parameters in Equations (1) and (2). This means that the smoothed shocks will also have contemporaneous correlation. As the expression for τ_t solves to be a weighted average of η_t beyond t, and the η_t are uncorrelated, τ_t is serially correlated, and hence the smoothed shocks will be as well. Their covariance matrix may, or may not, be singular.

4 Some Examples

4.1 A Basic Unobserved-Components Model

To illustrate the consequences of an excess of shocks in a stationary SSF, we first consider a simple example where the DGP for z_t, the observed variable, is $z_t = y_t$, with $y_t = u_t + v_t$, that is, the states are not autocorrelated. Let each of the uncorrelated model shocks have a known variance of unity so there are no parameters to estimate and therefore no identification issues.

The SSF of this DGP is $\varepsilon_t = \begin{bmatrix} u_t \\ v_t \end{bmatrix}$, $\psi_t = \begin{bmatrix} u_t \\ v_t \end{bmatrix}$, $M = 0, D_1 = \begin{bmatrix} 1 & 1 \end{bmatrix}$, $C = I$, and $R = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. The filtered shocks can be obtained using Equations (3)-(7), yielding $K_t = \begin{bmatrix} .5 \\ .5 \end{bmatrix}$, and hence $\hat{v}_t = \hat{u}_t = .5y_t$ as $z_t = y_t$. Accordingly, the estimated filtered shocks must be *perfectly correlated even though the model shocks (i.e. those in the DGP) are* *uncorrelated.*[3]

Turning to the impulse responses, the response of y_t to a one standard deviation shock in u_t in the model is unity. Now the standard deviation of the estimated shock is $\hat{u}_t = \frac{1}{\sqrt{.5}}$, as the perfect correlation of the empirical shocks implies $y_t = 2\hat{u}_t$. Consequently, the response of y_t to a one standard deviation shock in $\hat{u}_t$ is $\sqrt{2}$, which is 40% higher than that to u_t. The problem is that the correlation of the filtered shocks means that one will no longer get the same impulse responses as to the model shocks.

Now consider a model

$$\begin{aligned} y_t &= y_{1t} + y_{2t} \\ y_{1t} &= \rho y_{1t-1} + \varepsilon_{1t} \\ y_{2t} &= \sigma\varepsilon_{2t}, \end{aligned}$$

where the innovations ε_{1t} and ε_{2t} are assumed to be $n.i.d.(0, I_2)$, i.e. this is the DGP. This is an unobserved components model. Since we assume that all parameters of the DGP are known, we can write the model as

$$z_t = (1 - \rho L)y_t = \varepsilon_{1t} + \sigma\varepsilon_{2t} - \rho\sigma\varepsilon_{2t-1}.$$

To estimate the two innovations ε_{1t} and ε_{2t} from the one observed variable z_t we put the model into its SSF. So ψ_t equals $\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$ and the matrices are $M = 0, D_1 = \begin{bmatrix} 1 & \sigma \end{bmatrix}, D_2 = \begin{bmatrix} 0 & -\rho\sigma \end{bmatrix}, C = I$, and $R = 0$.

Using Equation (3), the estimated filtered shocks evolve as

$$E_t\psi_t = -K_t D_2 E_{t-1}\psi_{t-1} + K_t z_t.$$

[3]If the variances of the shocks were not unity K would have different elements, but that does not change the outcome that each measured shock is proportional to z_t.

Equation (4) gives the Kalman filter gain as

$$K_t = D_1'[D_2 P_{t-1|t-1} D_2' + D_1 D_1']^{-1},$$

and

$$(D_2 P_{t-1|t-1} D_2' + D_1 D_1')^{-1} = (\rho^2\sigma^2 P_{22,t-1|t-1} + 1 + \sigma^2)^{-1}.$$

So this is a scalar, c_t, and it means that $K_t = \begin{pmatrix} c_t \\ \sigma c_t \end{pmatrix}$. Consequently, $\Phi = -K_t D_2 = \begin{bmatrix} 0 & c_t\sigma\rho \\ 0 & c_t\sigma^2\rho \end{bmatrix}$ and $E_t\psi_t$ evolves as

$$\begin{aligned} E_t\psi_{1t} &= c_t\sigma\rho E_{t-1}\psi_{2t-1} + c_t z_t \\ E_t\psi_{2t} &= c_t\sigma^2\rho E_{t-1}\psi_{2t-1} + c_t\sigma z_t, \end{aligned}$$

giving $E_t\psi_{1t} = \sigma E_t\psi_{2t}$. This implies that the estimated filtered innovations $E_t\varepsilon_{1t}$ and $E_t\varepsilon_{2t}$ are *perfectly correlated*, something that is not true of the DGP innovations ε_{1t} and ε_{2t}. Later we will see that the same result emerges in many UC models with a single observed variable but more shocks, and the proof is the same as above.

It is also the case that there is serial correlation in the filtered innovations. Recall ψ_t equals $\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$ and in the steady-state of the expression for their filtered estimates, Equation (3), the coefficient on the first lag is $\Phi = -KD_2$, where K is the steady-state gain of the Kalman filter.[4] When $\sigma = 1$ and $\rho = .9$ then this gives $\Phi = \begin{bmatrix} 0 & .36 \\ 0 & .36 \end{bmatrix}$. Again the innovations of the DGP have no serial correlation. It is the inability to separate the innovations $\varepsilon_{1t|t}$ and $\varepsilon_{2t|t}$ when there is only one observed variable which results in serial correlation in the filtered shock innovations.

[4]This was also noted by Harvey (1992) for this type of model.

4.2 A Permanent/Transitory Components Model

We now consider the case where there are possibly $I(1)$ variables. In this instance it is useful to adopt the modified SSF described in the previous section. The modification is to the traditional observation equation and involves adding a term $D_2\psi_{t-1}$. This is useful because the observed non-stationary variables will be expressed as growth rates while the model variables will have been normalized by another I(1) variable to make them $I(0)$, i.e. they have been stationized.

To give a concrete example of a variable being stationized, in a DSGE model if the production function is $X_t = A_t N_t$, where X_t is output, A_t a permanent technology shock and N_t labour (hours), then the model will be expressed in terms of $I(0)$ variables by using stationized output, $\frac{X_t}{A_t}$. It is the log of this variable that will appear in the model as ψ_t, i.e. $\psi_t = x_t - a_t$ (lowercase denoting logs). These stationized model variables need to be related to observables, which are growth rates for the $I(1)$ variables, i.e. $z_t = \Delta x_t$. Hence $\Delta x_t = \Delta\psi_t + a_t$, and this adds a lagged state ψ_{t-1} to the observation equation.

To illustrate the consequences of excess shocks in models with I(1) variables we use a simple UC model that has been used to measure output gaps; see, for example, Orphanides and van Norden (2002). Such models decompose a series y_t as $y_t = y_t^p + y_t^c$, where y_t^p is a permanent or "trend" component of y_t and y_t^c a transitory component, often called the cycle (or output gap). Assumptions have to be made about how these evolve, and we look at the simplest set:

$$
\begin{aligned}
\psi_{1t} &= \Delta y_t^p = \varepsilon_{1t} \\
\psi_{2t} &= y_t^c = \varepsilon_{2t} \\
z_t &= \Delta y_t = \varepsilon_{1t} + \Delta\varepsilon_{2t},
\end{aligned}
$$

where ε_{1t} and ε_{2t} are $n.i.d.(0,1)$ and independent of one another. This model implies that Δy_t is a MA(1), $\Delta y_t = (1+\alpha L)u_t$, and so there are two parameters that can be estimated from the data $-$ α and σ_u^2 $-$ to give estimates of the variances of ε_{1t} and ε_{2t}. However, we will assume that all of the variance parameters are known so that no issues of identification arise.

To see the consequences of excess shocks, notice that this is the same model as analyzed in the previous sub-section, but now with $\rho = 1$, so the conclusions about singularity of the covariance matrix of the filtered shock innovations and serial correlation in the smoothed shock innovations still hold.

The perfect correlation between the estimated filtered shocks has economic implications. Note that in the simple UC model above the cyclical component - the output gap - is ε_{2t}. Consequently the estimated output gap and innovations to trend growth are perfectly correlated, making distinctions between aggregate demand and supply in the data difficult, even if they do actually exist in the DGP.

Morley et al. (2003) compared the Beveridge-Nelson (BN) definition of the cycle with that which one would obtain from an UC model. Drawing on Watson (1986), and using an UC model with uncorrelated innovations as th DGP, nesting the model above, they showed perfect correlation of the filtered shocks. They did this by noting that the BN decomposition led to a filtered estimate of the cycle, $E_t y_t^c$, identical to that from the Kalman filter, and it was already known that the BN trend and cycle innovations were perfectly correlated. They also show that although a BN decomposition based on an ARIMA model of the observed data implies a different UC model (and trend estimates), the estimated innovations to the trend and cycle components of this latter model are also perfectly correlated.[5] Our contribution is to point out that the perfect correlation arises from excess shocks, and so it is a wider problem than just filtering.

The DGP for the shocks can be changed so that ε_{1t} and ε_{2t} are correlated, e.g. as in the

[5] Anderson et al. (2006) demonstrate how the BN decomposition can be obtained from a "single source of error" state-space model (i.e. where the measurement and state equations are driven by a common innovation; see Snyder 1985). This utilizes the perfect correlation between the innovations. Morley (2002) provides an alternative state-space approach to computing the BN decomposition, which was adopted in Morley et al. (2003).

ARIMA model of z_t in Morley et al. (2003). It is then

$$\begin{aligned} y_t &= y_{1t} + y_{2t} \\ \Delta y_{1t} &= \varepsilon_{1t} \\ y_{2t} &= \varepsilon_{2t}, \end{aligned}$$

and the correlation between the DGP innovations can be captured with $\varepsilon_{1t} = \rho\varepsilon_{2t} + v_{1t}$, where v_{1t} and ε_{2t} are uncorrelated and have unit variances. This means

$$\begin{aligned} z_t &= \Delta y_t = \varepsilon_{1t} + \varepsilon_{2t} - \varepsilon_{2t-1} \\ &= \rho\varepsilon_{2t} + v_{1t} + \varepsilon_{2t} - \varepsilon_{2t-1}. \end{aligned}$$

Setting $\psi_t = \begin{bmatrix} v_{1t} \\ \varepsilon_{2t} \end{bmatrix}$ produces the matrices of the SSF

$$\begin{aligned} M &= 0, D_1 = \begin{bmatrix} 1 & 1+\rho \end{bmatrix}, \\ R &= 0, D_2 = \begin{bmatrix} 0 & -1 \end{bmatrix}, C = I. \end{aligned}$$

Solving as in the simple UC case of the previous sub-section we find that

$$E_t\psi_{2t} = (1+\rho)E_t\psi_{1t},$$

so the perfect correlation between filtered innovations holds regardless of the correlation between the DGP innovation processes.[6] Morley et al. (2003) note that Wallis (1995) made the point that the correlations between the DGP innovations and those evident in the estimated innovations can differ because of estimation. This agrees with our conclusion, as filtering is estimation of the states.

[6] The UC model of Clark (1987) has the same feature.

4.3 The Hodrick-Prescott Filter

The HP filter can be cast as an UC model (see, for example, Harvey and Jaeger 1993). Doing so enables the analysis of Section 3 to be readily applied to the HP filter. The corresponding UC model, which we use as the DGP, is

$$\begin{aligned}
(1-L)^2 y_t^p &= \varepsilon_{1t} \\
y_t^c &= \phi \varepsilon_{2t} \\
y_t &= y_t^p + y_t^c.
\end{aligned}$$

Note the cyclical component is a multiple (ϕ) of the innovation ε_{2t}. The parameter ϕ equals $\sqrt{\lambda}$ and λ =1,600 is common when the HP filter is applied to quarterly data.

The system can be expressed as

$$z_t = (1-L)^2 y_t = \varepsilon_{1t} + \phi(\varepsilon_{2t} - 2\varepsilon_{2t-1} + \varepsilon_{2t-2}).$$

Defining $\psi_t = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{2t-1} \end{bmatrix}$ the SSF has the matrices

$$\begin{aligned}
M &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \\
D_1 &= \begin{bmatrix} 1 & \phi & 0 \end{bmatrix}, D_2 = \begin{bmatrix} 0 & -2\phi & \phi \end{bmatrix}.
\end{aligned}$$

This leads to filtered innovations $E_t\varepsilon_{2t} = \phi E_t \varepsilon_{1t}$, so the filtered trend and cycle innovations are perfectly correlated. Of course these are shocks in the *one-sided HP filter*.

The original HP filter was two sided and therefore generated by smoothed, rather than filtered, shocks. In Section 3 it was shown that they will have serial correlation even if the filtered shocks do not. That is, even if the components model used to generate the HP filter above is the DGP there will *still be serial correlation in the smoothed cycle innovations.* This feature has often been observed in the literature for estimates of the cycle implied by the HP filter but has been interpreted as implying that the assumed UC model is mis-specified. However, here the serial correlation arises *even if the components model used to compute the HP filter is the DGP and it has innovations with no serial correlation. The serial correlation in the cyclical component obtained from the HP filter originates from excess shocks and the problems of extracting them in that case.*

As an experiment we simulated data using the model above as the DGP with $\phi = 40$, i.e. λ =1,600. Consequently the HP filter with a choice of λ =1,600 is the optimal filter. As predicted, applying it we find that innovations are perfectly correlated with $E_t\varepsilon_{2t} = 40E_t\varepsilon_{1t}$. There is little serial correlation in the filtered cycle innovation but large amounts in the smoothed one. As mentioned above, when this has been seen in data (where $E_T\varepsilon_{2t}$ is the extracted cyclical gap), it has often been said it shows that the trend has not been correctly estimated, and so the cyclical component has to be purged of its serial correlation. But here the model being used *is the DGP* and yet there is a large amount of persistence in the smoothed innovations.

More generally, the class of "structural trend/cycle" models set out by Harvey and Jaeger (1993) takes the form

$$\begin{aligned} y_t &= \mu_t + y_t^c && (13) \\ \Delta\mu_t &= \beta_{t-1} + \sigma_1\varepsilon_{1t} && (14) \\ \Delta\beta_t &= \sigma_2\varepsilon_{2t} && (15) \\ y_t^c &= \rho\cos\lambda y_{t-1}^c + \sin\lambda\psi_{t-1}^* + \sigma_3\varepsilon_{3t} && (16) \\ \psi_t^* &= -\rho\sin\lambda y_{t-1}^c + \cos\lambda\psi_{t-1}^* + \sigma_3\varepsilon_{4t}, && (17) \end{aligned}$$

where the ε_t are $N(0, I_4)$. This gives

$$\Delta^2 y_t = \varepsilon_{2t-1} + \Delta\varepsilon_{1t} + \Delta^2 y_t^C. \tag{18}$$

Harvey (1995) showed that the DGP of y_t^c implies

$$B(L)y_t^c = (1 - \rho\cos\lambda\, L\,)\,\sigma_3\varepsilon_{3t} + (\rho\sin\,\lambda\, L)\sigma_3\varepsilon_{4t}, \tag{19}$$

where $B(L) = 1 - 2\rho\cos\lambda L + \rho^2 L^2$. Hence the DGP for $\Delta^2 y_t$ can be written as

$$\begin{aligned} B(L)\Delta^2 y_t &= B(L)\varepsilon_{2t-1} + B(L)\Delta\varepsilon_{1t} + \\ &\quad \Delta^2[(1 - \rho\cos\lambda\, L\,)\,\sigma_3\varepsilon_{3t} + (\rho\sin\,\lambda\, L)\sigma_3\varepsilon_{4t}]. \end{aligned}$$

In general the covariance matrix of the filtered shocks will be singular when y_t is scalar. In Harvey and Jaeger's application of the model to US GDP $\sigma_1 = 0$, so that $\psi_t = \begin{bmatrix} \varepsilon_{2t} \\ \varepsilon_{3t} \\ \varepsilon_{4t} \end{bmatrix}$ and, repeating the derivations above, $E_t\varepsilon_{2t}$ and $E\varepsilon_{3t}$ are perfectly correlated.

4.4 An Estimated New-Keynesian Model with Excess Shocks

In this section we use an estimated New-Keynesian model from the literature, namely Ireland (2011), to examine the magnitudes of the correlations among the innovations that excess shocks can generate in a DSGE model.

Ireland (2011) uses a small-scale New-Keynesian (NK) model to compare the shocks driving the Great Recession to those of the previous two recessions, and to understand the slow recovery and the role of the zero lower bound on nominal interest rates. His model has more shocks than observables. A central aspect of his analysis is using the estimated shocks to track the recovery, so one needs them to be uncorrelated in order to say which shock is responsible. As we have said in previous sections this will not be possible. Ireland also presents variance decompositions assuming that the estimated shocks are uncorrelated, so that exercise is not using the properties of the data-based shocks, but simply the hypothetical ones in his model.

The model consists of three key equations: an IS equation, a NK Phillips Curve, and a Taylor rule. There are four shocks: preference a_t, technology Z_t, cost push $\hat{e}_t$ and monetary policy ε_{rt}. The technology shock is permanent, while the remainder are transitory. For completeness, the model equations are:

$$\hat{a}_t = \rho_a \hat{a}_{t-1} + \epsilon_{at}, \tag{20}$$

$$(z-\beta\gamma)(z-\gamma)\hat{\lambda}_t = \gamma z \hat{y}_{t-1} - (z^2+\beta\gamma^2)\hat{y}_t + \beta\gamma z \mathbb{E}_t(\hat{y}_{t+1}) + (z-\beta\gamma\rho_a)(z-\gamma)\hat{a}_t - \gamma z \hat{z}_t, \tag{21}$$

$$\hat{\lambda}_t = \hat{r}_t + \mathbb{E}_t(\hat{\lambda}_{t+1}) - \mathbb{E}_t(\hat{\pi}_{t+1}), \tag{22}$$

$$\hat{e}_t = \rho_e \hat{e}_{t-1} + \varepsilon_{et}, \tag{23}$$

$$\hat{z}_t = \varepsilon_{zt}, \tag{24}$$

$$(1+\beta\alpha)\hat{\pi}_t = \alpha\hat{\pi}_{t-1} + \beta\mathbb{E}_t(\hat{\pi}_{t+1}) - \psi\hat{\lambda}_t + \psi\hat{a}_t + \hat{e}_t, \tag{25}$$

$$\hat{r}_t - \hat{r}_{t-1} = \rho_\pi \hat{\pi}_t + \rho_g \hat{g}_t + \varepsilon_{rt}, \tag{26}$$

$$\hat{g}_t = \hat{y}_t - \hat{y}_{t-1} + \hat{z}_t, \tag{27}$$

$$0 = \gamma z \hat{q}_{t-1} - (z^2+\beta\gamma^2)\hat{q}_t + \beta\gamma z \mathbb{E}_t(\hat{q}_{t+1}) + \beta\gamma(z-\gamma)(1-\rho_a)\hat{a}_t - \gamma z \hat{z}_t, \tag{28}$$

and

$$\hat{x}_t = \hat{y}_t - \hat{q}_t. \tag{29}$$

The variables in the model are a Lagrange multiplier coming from the consumer's budget constraint λ_t ; output y_t; the growth rate of technology z_t; inflation π_t; interest rate r_t; output growth g_t; the efficient level of output q_t and the output gap x_t. Some of these variables have been normalized by the non-stationary technology Z_t so as to have a well-defined path to log-linearize about. $\hat{}$ denotes a log deviation from steady-state.

The parameters are the autocorrelation coefficients for the respective shock processes (ρ_a ρ_e); the steady-state growth rate of technology z; a discount factor β; an internal habits intensity γ; the degree of forward-looking behavior in price setting α; and interest rate rule parameters on inflation and growth (ρ_π and ρ_g).

Equations (20), (23) and (24) represent three of the model shock processes and the fourth shock ε_{rt} is that for monetary policy. The innovations $\varepsilon_{rt}, \varepsilon_{zt}$ etc., are all assumed to be uncorrelated and to have no serial correlation. Equations (21) and (22) together produce the NK IS curve. Equation (25) is a NK Phillips Curve. Equation (26) is the interest rate rule. Finally, Equation (29) is the output gap, with the natural rate of output defined by Equation (28).

Ireland estimated the model using maximum likelihood with **three** observed variables: output growth, inflation and the interest rate, using the sample 1983:1 to 2009:4. *As there are four model shocks, the model has more shocks than observed variables.*

We begin by replicating Ireland's parameter estimates exactly; the Maximum Likelihood Estimates (MLE) are presented in the left-hand side of Table 1.[7] As he found, the estimates of the two parameters - the degree of backward-looking behavior in price setting, α, and the degree of autocorrelation in the cost-push shock, ρ_e, - are at the boundary, and consequently we impose those values. In later variants, however, we will relax the restriction on the autocorrelation of the cost-push shock.

Table 2 reports the correlations among the estimated shock innovations, both filtered and smoothed. As predicted the filtered shocks have a singular covariance matrix, but this is not the case for smoothed shocks. It is apparent that there are some sizable bivariate correlations - for example, focusing on the filtered shocks, the correlation between the cost-push and technology shock is -0.64; for the cost-push and preference shock it is 0.46. The former

[7] The standard errors, however, are slightly different as Ireland bootstraps the model, whereas we do not.

Table 1: Parameter Estimates: Maximum Likelihood

Parameter	Replication Estimate	Excluding Preference Shock Estimate	Adding Hours Worked Estimate
γ	0.390	0.942	0.647
	(0.077)	(0.042)	(0.054)
ρ_π	0.415	0.419	0.439
	(0.046)	(0.044)	0.050
ρ_g	0.127	0.015	0.043
	(0.025)	(0.013)	(0.022)
ρ_a	0.980	0.000	0.861
	(0.025)	—	0.043
ρ_e	0.000	0.927	0.861
	—	(0.030)	(0.043)
Standard deviations of the shocks			
σ_a	0.087	0.000	0.033
	(0.102)	—	0.005
σ_e	0.002	0.004	0.009
	(0.0002)	(0.001)	0.001
$\sigma_{\hat{z}}$	0.010	0.110	0.005
	(0.002)	(0.077)	(0.0004)
σ_r	0.001	0.001	0.001
	(0.0001)	(0.0001)	(0.0001)

Standard errors are shown in parentheses.

correlation persists into the smoothed shocks, becoming -.56. So the shocks that Ireland estimates are mongrel shocks and, in particular, the contribution of empirical technology and cost push shocks cannot be separated. Consequently, an impulse response of an observed variable to a technology shock obtained from this model is a hypothetical experiment; in the estimated shocks it would have to be accompanied by an impulse to the cost-push shock.

In order to make the number of shocks equal to the number of observed variables either a shock could be omitted or an observed variable added. We do both. In the first case, we omit the preference shock. To be clear, this does *not* reflect a belief that it is unimportant, but is a way to examine the implications of eliminating the excess of shocks. In the second case, we add hours worked.[8] These result in several changes to the parameter estimates, in particular that the intensity of internal habits increases (γ), and the autocorrelation of the

[8]Adding hours worked was done by Pagan and Wickens (2019) for a related model, Ireland (2004). Hours worked can be substituted out of the model equations since the production function is linear in observed hours and so these can be mapped to stationized output.

Table 2: Properties of the Estimated Shock Innovations

	Replication				Excluding Preference Shock				Adding Hours Worked			
Shock innovation correlations												
Filtered												
	ε_a	ε_e	$\varepsilon_{\hat{z}}$	ε_r	ε_a	ε_e	$\varepsilon_{\hat{z}}$	ε_r	ε_a	ε_e	$\varepsilon_{\hat{z}}$	ε_r
ε_a	1	0.46	0.38	0.09	—	—	—	—	1	-0.27	0.29	0.14
ε_e	0.46	1	-0.64	0.22	—	1	0.09	0.12	-0.27	1	-0.09	-0.40
$\varepsilon_{\hat{z}}$	0.38	-0.64	1	-0.15	—	0.09	1	0.15	0.29	-0.09	1	0.05
ε_r	0.09	0.22	-0.15	1	—	0.12	0.15	1	0.14	-0.40	0.05	1
Smoothed												
ε_a	1	0.15	0.36	0.13	—	—	—	—	1	-0.27	0.29	0.14
ε_e	0.15	1	-0.56	0.00	—	1	0.09	0.12	-0.27	1	-0.09	-0.40
$\varepsilon_{\hat{z}}$	0.36	-0.56	1	-0.20	—	0.09	1	0.15	0.29	-0.09	1	0.05
ε_r	0.13	0.00	-0.20	1	—	0.12	0.15	1	0.14	-0.40	0.05	1
Shock innovation first-order autocorrelations												
	Filtered		Smoothed		Filtered		Smoothed		Filtered		Smoothed	
ε_a	0.32		0.39		—		—		0.02		0.02	
ε_e	-0.14		-0.53		-0.09		-0.09		0.81		0.81	
$\varepsilon_{\hat{z}}$	-0.15		-0.05		-0.46		-0.46		-0.03		-0.03	
ε_r	0.55		0.55		0.59		0.59		0.58		0.58	

cost-push shock increases (ρ_e) (see Table 1).

Turning to the correlation implications for the estimated innovations to the shocks when there are no excess shocks, they are generally considerably lower when the preference shock is excluded (Table 2). Focusing on this variant of the model, a particularly interesting case is the correlation between the cost-push and technology shocks, which drops down in absolute value from -0.64 to 0.09. The drop in the correlations is less pronounced when hours worked are added as an observed variable. It is noteworthy that in both variants even when there are no excess shocks there are non-zero estimated correlations. One explanation as to why this can occur is presented by Liu et al. (2018), namely the fact that DSGE models typically are over-identified. Another explanation, of course, is possible mis-specification of the model.

In the replication of Ireland's estimates both the filtered and smoothed shocks display considerable autocorrelation (bottom panel of Table 2). For example, the smoothed cost-push and monetary policy innovations have first-order autocorrelation coefficients of -0.53 and 0.55. When we omit the preference shock or add hours worked it is still present (e.g. in the monetary policy shock), suggesting the model may be mis-specified.

Consequently, it is useful to abstract from model mis-specification so as to isolate the implications of excess shocks. To do this a Monte Carlo experiment is conducted. More precisely, we generate samples of 108 observations (as in Ireland 2011) by simulating the model with *independent* uncorrelated shock innovations, and then estimate the model using maximum likelihood and the simulated data. This is done repeatedly until we obtain $1,000$ estimates where the second-derivative matrix of the log likelihood at the optimum was negative definite. *The model used for simulation is the DGP - there is no mis-specification.* We produce two sets of estimates - one when Ireland (2011) is the DGP and therefore there are excess shocks, and the other when the DGP omits the preference shock.

Table 3: Shock Innovation Correlations: Monte Carlo Analysis With and Without Excess Shocks

	With Excess Shocks				Without Excess Shocks			
				Filtered				
	ε_a	ε_e	$\varepsilon_{\hat{z}}$	ε_r	ε_a	ε_e	$\varepsilon_{\hat{z}}$	ε_r
ε_a	1	0.54	0.32	-0.01	—	—	—	—
		[0.39,0.67]	[0.16,0.47]	[-.07,0.06]				
ε_e	0.54	1	-0.62	0	—	1	0	-0.01
	[0.39,0.67]		[-0.74,-0.48]	[-0.08,0.08]			[-0.07,0.08]	[-0.09,0.07]
$\varepsilon_{\hat{z}}$	0.32	-0.62	1	0	—	0	1	0.01
	[0.16,0.47]	[-.74,-0.48]		[-.09,0.08]		[-0.07,0.08]		[-0.14,0.16]
ε_r	-0.01	0	0	1	—	-0.01	0.01	1
	[-0.07,0.06]	[-0.08,0.08]	[-0.09,0.08]			[-0.09,0.07]	[-0.14,0.16]	
				Smoothed				
ε_a	1	0.33	0.28	-0.01	—	—	—	—
		[0.20,0.47]	[0.13,0.44]	[-0.07,0.06]				
ε_e	0.33	1	-0.39	0	—	1	0	-0.01
	[0.20,0.47]		[-0.51,-0.26]	[-0.11,12]			[-0.07,0.08]	[-0.09,0.07]
$\varepsilon_{\hat{z}}$	0.28	-0.38	1	0	—	0	1	0.01
	[0.13,0.44]	[-0.51,-0.26]		[-0.10,0.09]		[-0.07,0.08]		[-0.14,0.16]
ε_r	-0.01	0	0	1	—	-0.01	0.01	1
	[-0.07,0.06]	[-0.11,0.12]	[-0.01,0.09]			[-0.09,0.07]	[-0.14,0.16]	

Notes: The table presents the mean estimate, together with the 5th and 95th percentiles (in brackets). "Without Excess Shocks" refers to the model excluding the preference shock. Sample size is 108 observations. 1,000 successful replications used.

The results are stark (Table 3). When there are excess shocks but no model mis-specification, so that the shock innovations are uncorrelated, the estimated shocks display sizable bivariate correlations. For example the mean estimate of the correlation between the cost-push and technology shocks is -0.62. Alternatively, in the absence of excess shocks it is zero. Such a dramatic drop is evident across many of the correlations.

The means of the first-order autocorrelations in the shock innovations from the Monte Carlo analysis are reported in Table 4. These autocorrelations are generally considerably smaller than those for shocks estimated from the actual data, which are reported in Table 2. This is true when there are no excess shocks and excess shocks. These results suggests that: firstly, in the instance of the Ireland (2011) model, most of the autocorrelation in the estimated shocks stems from model mis-specfication; and secondly that the main consequence of having excess shocks in this instance is the generation of sizable bivariate correlations among them.

Table 4: Shock Innovation First-order Autocorrelations: Monte Carlo Analysis With and Without Excess Shocks

	With Excess Shocks		Without Excess Shocks	
	Filtered	Smoothed	Filtered	Smoothed
ε_a	0	0.087	—	—
	[-0.137,0.139]	[-0.058,0.226]	—	—
ε_e	-0.003	-0.347	-0.006	-0.006
	[-0.123,0.115]	[-0.477,-0.210]	[-0.159,143]	[-0.159,143]
$\varepsilon_{\hat{z}}$	-0.008	0.110	-0.008	-0.008
	[-0.120,0.110]	[-0.021,0.236]	[-0.156,0.146]	[-0.156,0.146]
ε_r	-0.016	-0.016	-0.020	-0.020
	[-0.174,0.132]	[-0.174,0.132]	[-0.183,0.138]	[-0.183,0.138]

Notes: The table presents the mean estimate, together with the 5th and 95th percentiles (in brackets). Sample size is 108 observations. 1,000 successful replications used. "Without Excess Shocks" refers to the model excluding the preference shock.

5 Some Questions That Might be Asked

5.1 Is This Just a Parameter Identification Problem?

No. It is a consequence of *partial information*. To emphasize this we consider here a very simple model, which is a particular case of the more general analysis presented in Section 2.

Suppose we have one observed variable, y_{1t}, which is a linear function with known parameters a_{11} and a_{12} of two shocks ε_{1t} and ε_{2t}. These are taken to be normally distributed, with known standard deviations σ_1 and σ_2. The specification then is

$$y_{1t} = a_{11}\varepsilon_{1t} + a_{12}\varepsilon_{2t}. \tag{30}$$

All parameters are known. There is no parameter identification problem. However, with only y_t observed, excess shocks exist and we *do not have enough information to uniquely estimate ε_{1t} and ε_{2t}* separately.

Suppose, instead we have a second observed variable, y_{2t}. This also is a linear function of the same shocks, with known parameters a_{21} and a_{22}:

$$y_{2t} = a_{21}\varepsilon_{1t} + a_{22}\varepsilon_{2t}. \tag{31}$$

Equations (30) and (31) can be summarized as a system of equations

$$y_t = A\varepsilon_t, \tag{32}$$

where $y_t \equiv (y_{1t}, y_{2t})'$, $\varepsilon_t \equiv (\varepsilon_{1t}, \varepsilon_{2t})'$ and $A \equiv \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$. A is known. As long as A is non-singular, because y_t is comprised of two observed variables, y_{1t} and y_{2t}, *we now have enough information* to uniquely estimate ε_{1t} and ε_{2t} separately.

5.2 Does Bayesian Estimation of the Parameters Provide a Solution?

No. Consider Bayesian estimation of the unobserved components model where some priors could be placed on the variances of the components or on ρ. This will result in some parameter estimates that will likely be different to the MLE in small samples (and also possibly the true values of the parameters in the DGP). However, as seen in our derivation, *for any given value of these parameters there is a singularity of the spectrum of filtered shocks* when there are excess shocks. This is also the case if the parameters are calibrated. If the Bayesian estimates differ from the MLE estimates there may be different bivariate correlations and serial correlations in the estimated innovations, but the correlations will still exist. As the sample size grows of course these differences will disappear as the priors get dominated. So what differences there are will be specific to the context but they will always feature correlated innovations when there are excess shocks regardless of which parameter estimates are used.

To examine the possible consequences in a DSGE context, we re-estimate Ireland's model (2011) using Bayesian methods. We assign priors to the estimated parameters using standard distributions - such as Beta distributions for the autocorrelation coefficients, and Inverse Gamma distributions for the standard deviations of the shock. We use random-walk Metropolis Hastings to simulate the posterior with two chains of 300,000 observations, dropping the first 75 per cent as burn-in.[9] The resulting parameter estimates, shown in Table 5, are qualitatively similar to those obtained with maximum likelihood that are reported in Table 1. Table 5 also presents estimates for the variants without excess shocks.

Again, the smoothed shock innovations are substantially correlated (Table 6). The strong negative correlation between the innovations to the technology and cost push shock remains, as does the first-order autocorrelation in many of the innovations. Making the number of shocks equal to the number of observed variables by eliminating the preference shock, or adding hours worked as an observed variable, once again generally reduces the bivariate correlations. Some substantial correlations, however, still exist, as does first-order autocorrelation, for example, in the monetary policy shock. That these correlations exist even in the absence of excess shocks suggests the model may be mis-specified.

Once again, in order to abstract from mis-specification we conduct Monte Carlo analysis,

[9] The acceptance rates for Ireland's specification are 35.9 and 35.3 per cent. We assess the chains convergence using methods based on Brooks and Gelman (1998).

Table 5: Parameter Estimates: Bayesian

Parameter	Prior	Replication Posterior Mean & 90% H. P. D.	Excluding Preference Shock Posterior Mean & 90% H. P. D.	Adding Hours Worked Posterior Mean & 90% H. P. D.
γ	B(0.5,0.2)	0.418	0.880	0.556
		(0.294-0.543)	(0.826-0.937)	(0.481-0.634)
ρ_π	G(0.5,0.2)	0.418	0.450	0.396
		(0.344-0.490)	(0.376-0.521)	(0.328-0.461)
ρ_g	G(0.1,0.05)	0.122	0.025	0.108
		(0.086-0.158)	(0.011-0.039)	(0.075-0.141)
ρ_a	B(0.5,0.2)	0.948	0.000	0.938
		(0.921-0.982)	—	(0.906-0.971)
ρ_e	B(0.5,0.2)	0.000	0.870	0.916
		—	(0.828-0.914)	(0.876-0.956)
Standard deviations of the shocks				
σ_a	IG(0.001,0.5)	0.045	0.000	0.042
		(0.023-0.066)	—	(0.025-0.060)
σ_e	IG(0.001,0.5)	0.002	0.003	0.002
		(0.001-0.002)	(0.003-0.004)	(0.001-0.002)
$\sigma_{\hat{z}}$	IG(0.001,0.5)	0.010	0.06	0.005
		(0.007-0.013)	(0.032-0.085)	(0.005-0.006)
σ_r	IG(0.001,0.005)	0.001	0.001	0.001
		(0.001-0.002)	(0.001-0.002)	(0.001-0.002)

Notes: Distributions: B - Beta; G - Gamma; IG - Inverse Gamma. $B(\mu, \sigma)$: μ is the mean and σ the standard deviation. H. P. D. is Highest Posterior Density interval.

using the posterior mean of Ireland's model with excess shocks as the DGP (Table 7).[10] The results are similar to those obtained with maximum likelihood presented in Table 3. It is apparent that sizable correlations persist and autocorrelation is also evident in the estimated cost-push shock innovations, even though the simulated innovations were independent. This reflects the presence of excess shocks, and so Bayesian estimation is not a solution, unless the priors effectively exclude a shock.

5.3 Do Measurement Errors Produce Excess Shocks?

Yes. To isolate the impact of including measurement error we focus on the variant of Ireland's model that has only three shocks, that is our baseline model has the same number of shocks

[10]We use a single chain of 200,000 observations for each sample.

Table 6: Properties of the Estimated Smoothed Shock Innovations: Bayesian

	Replication				Excluding Preference Shock				Adding Hours Worked			
Shock innovation correlations												
	ε_a	ε_e	$\varepsilon_{\hat{z}}$	ε_r	ε_a	ε_e	$\varepsilon_{\hat{z}}$	ε_r	ε_a	ε_e	$\varepsilon_{\hat{z}}$	ε_r
ε_a	1	0.16	0.31	0.08	—	—	—	—	1	-0.01	0.09	0.06
ε_e	0.16	1	-0.60	-0.01	—	1	0.18	-0.01	0.16	1	-0.27	0.08
$\varepsilon_{\hat{z}}$	0.31	-0.60	1	-0.20	—	0.18	1	0.31	0.09	-0.27	1	-0.13
ε_r	0.08	-0.01	-0.20	1	—	-0.01	0.15	1	0.06	0.08	-0.13	1
Shock innovation first-order autocorrelations												
ε_a	0.38				—				0.17			
ε_e	-0.56				-0.01				-0.47			
$\varepsilon_{\hat{z}}$	-0.08				-0.41				-0.05			
ε_r	0.55				0.59				0.55			

Estimates are at the mean of the posterior.

as observed variables. We then assume that observed output growth is measured with error, thereby producing a standard form of the corresponding measurement equation:

$$\Delta y_t^{obs} = \hat{g}_t + \varepsilon_{\Delta y_t}. \tag{33}$$

Here Δy_t^{obs} is the observed output growth and $\varepsilon_{\Delta y_t}$ is the measurement error. This formulation, while common, does not maintain co-integration between the model's output variable and the observed data - see Pagan and Robinson (2019). Nevertheless using this with the model which excludes the preference shock yields the parameter estimates reported in Table 8. The most noteworthy difference is that the aggressiveness of the monetary policy response to output growth (ρ_g), increases.

The consequences of including measurement error for the measured shock innovation correlations are shown in Table 9. Comparing the left-hand column (without measurement error) and the right-hand (with) it is apparent that its inclusion results in much more sizable correlations. These are not only between the measurement error and the model (structural) shocks - for example, the correlation with the filtered monetary policy shock is 0.43 - but also between model shocks themselves. The correlation of the innovation of technology growth and the monetary policy shock is particularly strong (-0.71).

Table 7: Shock Innovation Properties: Monte Carlo Analysis With Bayesian Estimation

	Smoothed			
	Shock innovation correlations			
	ε_a	ε_e	$\varepsilon_{\hat{z}}$	ε_r
ε_a	1	0.33	0.20	-0.01
		[0.20,0.48]	[0.05,0.37]	[-0.06,0.03]
ε_e	0.33	1	-0.47	0
	[0.20,0.48]		[-0.57,-0.38]	[-0.10,0.09]
$\varepsilon_{\hat{z}}$	0.20	-0.47	1	0
	[0.05,0.37]	[-0.57,-0.38]		[-0.08,0.06]
ε_r	-0.01	0	0	1
	[-0.06,0.03]	[-0.10,0.09]	[-0.07,0.06]	
	Shock innovation first-order autocorrelations			
ε_a	0.08			
	[-0.02,0.19]			
ε_e	-0.37			
	[-0.48,-0.25]			
$\varepsilon_{\hat{z}}$	0.10			
	[-0.02,0.21]			
ε_r	-0.01			
	[-0.14,0.11]			

Notes: The table presents the mean estimate, together with the 5th and 95th percentiles (in brackets). Sample size is 108 observations. 100 successful replications used.

In contrast, the estimated shocks in the model including measurement error tend to have less first-order autocorrelation than when measurement error is excluded. This suggests that these autocorrelations are being driven by a mis-specified model, consistent with the findings in Sub-section 4.4.

To isolate the impact of including measurement error we abstract from model mis-specification by again using a Monte Carlo experiment (Table 10). *Excess shocks stemming from measurement error result in substantial bivariate correlations in the estimated filtered model shocks.* An example of this is that the cost-push and technology shock innovations have a mean correlation of -0.75. This occurs despite measurement error actually being present in the DGP.

It is possible to include measurement error and not have excess shocks if the model has *less* structural shocks than observed variables. This, however, means that whenever data without measurement error were to become available, then the model would be stochastically singular

Table 8: Parameter Estimates: Maximum Likelihood with Measurement Error

Parameter	Excluding Preference Shock Estimate	Excluding Preference Shock and Including Measurement Error Estimate
γ	0.942	0.996
	(0.042)	(0.003)
ρ_π	0.419	0.673
	(0.046)	0.108
ρ_g	0.015	1.26
	(0.013)	(0.772)
ρ_e	0.927	0.933
	(0.030)	(0.029)
Standard deviations of the shocks		
σ_e	0.004	0.004
	(0.001)	0.001
$\sigma_{\hat{z}}$	0.110	0.072
	(0.077)	(0.046)
σ_r	0.001	0.001
	(0.0001)	(0.0001)
$\sigma_{\Delta y}$	—	0.007
		(0.0004)

Standard errors are shown in parentheses.

and therefore could not be estimated with standard likelihood-based methods, although other methods always exist (e.g. Canova et al. 2014).

In summary, while measurement error is widely used in estimated DSGE models and may be intuitively appealing, it can result in sizable bivariate correlations between the estimated shock innovations, impeding their interpretation.

5.4 Are the Results Due to a Small Sample Size?

No. Ireland (2011) is estimated over a short sample, namely 1983:1 to 2009:4 (108 observations). One concern might be that the sizable correlations among the estimated shocks reflect this short sample, and would become unimportant with longer-run data. To examine this we once again use Monte Carlo analysis. We produce one set of estimates with 108 observations - the same as Ireland (2011) - and another with considerably more (1,000 observations).

We expect that the filtered shocks will have a singular covariance matrix independent of the sample size, and this is confirmed. The results in Table 11 confirm that the sizable cor-

Table 9: Properties of the Estimated Shock Innovations: Maximum Likelihood with Measurement Error

	Excluding Measurement Error				Including Measurement Error			
Shock innovation correlations								
Filtered								
	ε_e	$\varepsilon_{\hat{z}}$	ε_r	$\varepsilon_{\Delta y}$	ε_e	$\varepsilon_{\hat{z}}$	ε_r	$\varepsilon_{\Delta y}$
ε_e	1	0.09	0.12	—	1	0.25	-0.03	0.14
$\varepsilon_{\hat{z}}$	0.09	1	0.15	—	0.25	1	-0.71	0.32
ε_r	0.12	0.15	1	—	-0.03	-0.71	1	0.43
$\varepsilon_{\Delta y}$	—	—	—	—	0.14	0.32	0.43	1
Smoothed								
ε_e	1	0.09	0.12	—	1	0.22	-0.05	0.15
$\varepsilon_{\hat{z}}$	0.09	1	0.15	—	0.22	1	-0.56	0.14
ε_r	0.12	0.15	1	—	-0.05	-0.56	1	0.29
$\varepsilon_{\Delta y}$	—	—	—	—	0.15	0.14	0.29	1

	Excluding Measurement Error		Including Measurement Error	
Shock innovation first-order autocorrelations				
	Filtered	Smoothed	Filtered	Smoothed
ε_e	-0.09	-0.09	0	0
$\varepsilon_{\hat{z}}$	-0.46	-0.46	0.01	0.02
ε_r	0.59	0.59	0	0
$\varepsilon_{\Delta y}$	—	—	0.02	0.02

Note: Excludes preference shock.

relations persist even when longer time-series of data are used in estimation than is typically available in macroeconomics. For example, with 1,000 observations the mean estimate of the correlations between the smoothed innovations to the filtered technology and cost-push shocks is -0.62, even though the DGP innovations are uncorrelated.

Table 10: Shock Innovation Correlations: Monte Carlo Analysis With Measurement Error

	Estimated Model and DGP Includes Measurement Error			
	Filtered			
	ε_e	$\varepsilon_{\hat{z}}$	ε_r	$\varepsilon_{\Delta y}$
ε_e	1	-0.75	0.32	0.04
		[-0.87,-0.61]	[0.12,0.53]	[-0.12,0.20]
$\varepsilon_{\hat{z}}$	-0.75	1	0.37	0.04
	[-0.87,0.61]		[0.25,0.48]	[-.13,0.21]
ε_r	0.32	0.37	1	-0.04
	[0.12,0.53]	[0.25,0.48]		[-0.19,0.12]
$\varepsilon_{\Delta y}$	0.04	0.04	-0.04	1
	[-0.12,0.20]	[-0.13,0.21]	[-0.19,0.12]	
	Smoothed			
ε_e	1	-0.61	0.18	0.02
		[-0.74,-0.46]	[0.05,0.34]	[-0.13,0.18]
$\varepsilon_{\hat{z}}$	-0.61	1	0.22	0.02
	[-0.74,-0.46]		[0.13,0.31]	[-0.13,0.19]
ε_r	0.18	0.22	1	-0.02
	[0.05,0.34]	[0.13,0.31]		[-0.18,0.14]
$\varepsilon_{\Delta y}$	0.02	0.02	-0.02	1
	[-0.13,0.18]	[-0.13,0.19]	[-0.18,0.14]	

Notes: the table presents the mean estimate, together with the 5th and 95th percentiles (in brackets). Excludes the preference shock. Sample size is 108 observations. 1,000 successful replications used.

5.5 Why Can we Have Excess Shocks with Factor Models?

A factor model is one case where there can be more shocks than observable variables and the empirical shocks may be uncorrelated. To see why, consider a factor model such as

$$y_{it} = y_t^* + \varepsilon_{it},$$

where y_{it} is an observed variable, y_t^* the latent factor and ε_{it} an idiosyncratic shock. Now there are excess shocks. We could estimate the factor and the idiosyncratic shocks as $\hat{y}_t^* = \frac{1}{n}\sum_{j=1}^{n} y_{it}$ and $\hat{\varepsilon}_{it} = y_{it} - \hat{y}_t^*$. The covariance between the estimated factor and the *ith*

Table 11: Shock Innovation Correlations: Monte Carlo Analysis Varying the Sample Size

	108 Observations				1,000 Observations			
	Filtered							
	ε_a	ε_e	$\varepsilon_{\hat{z}}$	ε_r	ε_a	ε_e	$\varepsilon_{\hat{z}}$	ε_r
ε_a	1	0.54	0.32	-0.01	1	0.54	0.33	0
		[0.39,0.67]	[0.16,0.47]	[-.07,0.06]		[0.49, 0.59]	[-.27,0.38]	[-0.02,0.2]
ε_e	0.54	1	-0.62	0	0.54	1	-0.62	0
	[0.39,0.67]		[-0.74,-0.48]	[-0.08,0.08]	[0.49,0.59]		[-0.66,-0.57]	[-0.03,0.03]
$\varepsilon_{\hat{z}}$	0.32	-0.62	1	0	0.33	-0.62	1	0
	[0.16,0.47]	[-.74,-0.48]		[-.09,0.08]	[0.27,0.38]	[-0.66,-0.57]		[-0.03,0.03]
ε_r	-0.01	0	0	1	0	0	0	1
	[-0.07,0.06]	[-0.08,0.08]	[-0.09,0.08]		[-0.02,0.02]	[-0.03,0.03]	[-0.03,0.03]	
	Smoothed							
ε_a	1	0.33	0.28	-0.01	1	0.59	0.38	0.02
		[0.20,0.47]	[0.13,0.44]	[-0.07,0.06]		[0.29,0.38]	[0.24,0.34]	[-0.02,0.02]
ε_e	0.33	1	-0.39	0	0.59	1	-0.57	0.03
	[0.20,0.47]		[-0.51,-0.26]	[-0.11,12]	[0.29,0.38]		[-0.42,-0.34]	[-0.04,0.04]
$\varepsilon_{\hat{z}}$	0.28	-0.38	1	0	0.38	-0.58	1	0.03
	[0.13,0.44]	[-0.51,-0.26]		[-0.10,0.09]	[0.24,0.34]	[-0.42,-0.34]		[-0.03,0.03]
ε_r	-0.01	0	0	1	0.02	0.03	0.03	1
	[-0.07,0.06]	[-0.11,0.12]	[-0.01,0.09]		[-0.02,0.02]	[-0.04,0.04]	[-0.03,0.03]	

Notes: The table presents the mean estimate, together with the 5th and 95th percentiles (in brackets). 1,000 successful replications used. Estimated with maximum likelihood.

idiosyncratic shock is

$$\begin{aligned} cov(\hat{y}_t^* \hat{\varepsilon}_{it}) &= \mathbb{E}[(\frac{1}{n}\sum_{j=1}^{n} y_{jt})(y_{it} - \frac{1}{n}\sum_{j=1}^{n} y_{jt})] \\ &= \mathbb{E}[(y_t^* + \frac{1}{n}\sum_{j=1}^{n} \varepsilon_{jt})(\varepsilon_{it} - \frac{1}{n}\sum_{j=1}^{n} \varepsilon_{jt})] \\ &= \mathbb{E}[\frac{1}{n}\sum_{j=1}^{n} \varepsilon_{jt})(\varepsilon_{it} - \frac{1}{n}\sum_{j=1}^{n} \varepsilon_{jt})] \\ &= \frac{\sigma_i^2}{n} - \frac{\sigma_i^2}{n^2} \end{aligned}$$

which tends to zero as $n \to \infty$, i.e. when the number of observables tends to infinity we can have more shocks than observables. Such is the way with infinity.

6 Conclusion

Estimating structural shocks is a key aspect of applied macroeconomics, and the consequences of not having enough information to do so were examined. It was shown that when there are excess shocks - the number of shocks exceeds the number of observed variables - the estimated structural shocks will be correlated, even when the true (DGP) shocks are not.

This is a problem of not having enough information, and can occur in models where all parameters are known and therefore identified. Excess shocks were demonstrated to occur in a wide variety of models used in macroeconomics. It is particularly important for models where estimated shocks are used for interpretation, for example, by analyzing impulse responses. DSGE models are a noteworthy case. Interpretations based on the assumed properties of the shocks, rather than those of the estimated shocks, can be misleading. When excess shocks exist an impulse to one shock is an experiment that may not be representative of the estimated shocks.

The quantitative importance of the consequences of excess shocks was examined with an estimated New-Keynesian DSGE model from the literature, Ireland (2011). The resulting correlations among the estimated shocks were shown to be sizable. This occurred regardless of whether either maximum likelihood or Bayesian estimation methods are used, and whether there are small or large samples. It was demonstrated that allowing for measurement error will frequently result in excess shocks and sizable correlations, casting doubt on the usefulness of that methodology.

The major implication of our analysis is that excess shocks should be avoided. Using more observed data in estimation or redesigning the model to have less shocks would do this. Remaining correlations between shocks may then reflect mis-specification and not be a consequence of trying to extract more information than is possible. At the very least when there are excess shocks the correlations between the estimated shock innovations should be reported. It may be the case that one of the shocks is uncorrelated with the others, even if they cannot all be. Otherwise, too many shocks spoil the interpretation.

References

Anderson, H. M., C. N. Low and R. Snyder (2006) "Single Source of Error State Space Approach to Beveridge Nelson Decomposition", *Economics Letters*, 91, 104-109.

Brooks, S. P. and Gelman (1998) "General Methods for Monitoring Convergence of Iterative Simulations", *Journal of Computational and Graphical Statistics*, 7, 3, 434-455.

Campbell S. L. and C. D. Meyer (2009) *Generalized Inverses of Linear Transformations*, Society for Industrial and Applied Mathematics, Philadelphia, USA.

Canova, F., F. Ferroni and C. Matthes (2014) "Choosing the Variables to Estimate Singular DSGE Models", *Journal of Applied Econometrics*, 29, 7, 1099-1117.

Canova, F. and F. Ferroni (2019) "Mind the Gap! Stylized Dynamic Facts and Structural Models", *CEPR Discussion Papers*, 13948.

Christiano, L. J., R. Motto and M. Rostagno (2014), "Risk Shocks", *American Economic Review*, 104, 1, 27-65.

Clark, P. K. (1987) "The Cyclical Component of U.S. Economic Activity", *The Quarterly Journal of Economics* 102, 797–814.

Harvey, A.C. (1985) "Trends and Cycles in Macroeconomic Time Series", *Journal of Business & Economic Statistics,* 3, 216-227.

Harvey, A. C. (1990) *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge UK.

Harvey, A. C. (1992) "On Trend Extraction Models: Comments on a Paper by Garcia-Ferrer and del Hoyo". *Journal of Forecasting*, 11, 669-672.

Harvey, A. C. and Jaeger, A. (1993) "Detrending, Stylized Facts and the Business Cycle", *Journal of Applied Econometrics*, 8, 3, 231-247.

Hodrick, R. and E. Prescott (1997) "Post-war Business Cycles: An Empirical Investigation", *Journal of Money, Credit and Banking*, 29, 1-16.

Ireland, P. N. (2004) "Technology Shocks in the New Keynesian Model", *Review of Economics and Statistics*, 86, 923-936.

Ireland, P. N. (2011) "A New Keynesian Perspective on the Great Recession", *Journal of Money, Credit and Banking*, 43, 31-54.

Kurz, M. S. (2018) "A Note on Low-dimensional Kalman Smoothers for Systems with Lagged States in the Measurement Equation", *Economic Letters*, 168, 42-45.

Liu, X., A. R. Pagan and T. Robinson (2018) "Critically Assessing Estimated DSGE Models: A Case Study of a Multi-sector Model", *Economic Record*, 94, 349-371.

Morley, J. C. (2002) "A State-space Approach to Calculating the Beveridge-Nelson Decomposition", *Economics Letters*, 75, 123-127.

Morley, J. C., C. R. Nelson and E. Zivot (2003) "Why are The Beveridge-Nelson and Unobserved Components Decompositions of GDP so Different?", *The Review of Economics and Statistics*, LXXXV, 235-243.

Nimark, K. P. (2015) "A Low Dimensional Kalman Filter for Systems With Lagged States in the Measurement Equation", *Economics Letters*, 127(C), 10-13.

Orphanides, A. and S. van Norden (2002), "The Unreliability of Output-Gap Estimates in Real Time", *Review of Economics and Statistics*, 84, 4, 569-583.

Pagan A. R. and T. Robinson (2019) "Implications of Partial Information for Applied Macroeconomic Modelling", *Melbourne Institute: Applied Economic and Social Research Working Paper*, University of Melbourne, 12/19.

Pagan A. R. and M. Wickens (2019) "Checking if the Straitjacket fits", *CAMA Working Papers*, Centre for Applied Macroeconomic Analysis, Australian National University, 2019-81.

Ravenna, F. (2007) "Vector Autoregressions and Reduced Form Representations of DSGE Models", *Journal of Monetary Economics*, 54, 7, 2048-2064.

Smets, F. and R. Wouters (2007) "Shocks and Frictions in US Business Cycles: a Bayesian DSGE Approach", *American Economic Review*, 97, 586-606.

Snyder, R. D. (1985) "Recursive Estimation of Dynamic Linear Models", *Journal of the Royal Statistical Society*, Series B, 47, 272-276.

Wallis, K. F. (1995) Discussion of "Trend Extraction: a Practitioner's Guide" by M.R. Wickens, Annex E, *Government Economic Service Working Paper*, London: HM Treasury, 125.

Watson, M. (1986) "Detrending with Stochastic Trends", *Journal of Monetary Economics*, 18, 49-75.

THE UNIVERSITY OF MELBOURNE

POSTERA CRESCAM LAUDE

MELBOURNE INSTITUTE