Confronting Theory with Data: Model Validation and DSGE Modeling

Niraj Poudyal and **Aris Spanos**
Department of Economics,
Virginia Tech, USA

April 2013

Abstract

The primary objective of this paper is to discuss the problem of confronting theory with data using a DSGE model as an example. The paper calls into question the traditional approach for securing the empirical validity of DSGE models on several grounds, including identification, statistical misspecification, substantive adequacy, poor forecasting ability and misleading policy analysis. It is argued that most of these weakness stem from failing to distinguish between statistical and substantive inadequacy and secure the former before assessing the latter. The paper disentangles the statistical from the substantive premises of inference with a view to unveil the above mentioned problems. The critical appraisal is based on a particular DSGE model using USA macro-data for the period 1948-2010. It is shown that this model is statistically misspecified and when respecified with a view to achieve statistical adequacy one needs to adopt the Student's t VAR model. The latter model is shown to provide a sound basis for forecasting and policy simulations, and can be used to guide the search for better DSGE models.

1 Introduction

The recent 2008 global financial crises that threatened the world's monetary system, and the reaction of the economics professsion on how to address the looming global recession, raised several questions pertaining to economics as a scientific discipline. In particular, the soundness of its empirical underpinnings. How do we acquire causal knowledge about economic phenomena? How do we distinguish between fact and fiction when interpreting economic data? How do we distinguish between well-grounded knowledge and speculation stemming from personal beliefs? How do we differentiate between 'good' theories and 'bad' theories? What is the role of the data in testing the adequacy of theories? What is the scope of empirical macroeconometric models in forecasting and policy analysis?

The primary aim of this paper is to propose reasoned answers to the above questions. Section 2 provides a brief historical introduction to macroeconometric modeling as a prelude to the discussion of different methodological perspectives that have influenced to a greater or a lesser extent the development of macroeconometric modeling during the 20th century. Section 3 brings out the untrustworthiness of evidence problem arising from estimating (quantifying) the structural model directly. This strategy often leads to an estimated model which is both statistically and substantively misspecified but one has no principled way to distinguish between the two and apportion blame. It is argued that the key to addressing this Duhemian ambiguity is the distinction between the substantive and statistical premises of inductive inference. The Simultaneous Equations Model (SEM) is used to both illustrate this distinction as well as bring out the problems raised by ignoring it. The error statistical perspective developed in sections 2-3 is applied to DSGE modeling in section 4. The substantive and statistical premises are distinguished so that one can secure the validity of the statistical premises before probing substantive questions.

2 Empirical modeling in economics

Statistical modeling was widely adopted in empirical modeling in economics in the early 20th century, but after a century of extensive empirical research the question 'when do data $\mathbf{Z}_0$ provide evidence *for* a particular hypothesis or theory?' has not been adequately answered. The current practice uses a variety of criteria for answering this question, including [i] statistical (goodness-of-fit/prediction), [ii] substantive (theoretical meaningfulness) and [iii] pragmatic (simplicity, elegance, mathematical rigor). Such criteria, however, are of questionable credibility when the estimated model does not account for the statistical regularities in the data; see Spanos (2007).

The reliability of empirical evidence in economics stems from two separate but related dimensions of empirical modeling. The first pertains to how well a theory model 'captures' (describes, explains, predicts) the key features of the phenomenon of interest, and is referred to as *substantive adequacy.* The second – often implicit – is concerned with the validity of the probabilistic assumptions imposed on the observable stochastic process $\{\mathbf{Z}_t,\ t{\in}\mathbb{N}{:=}(1,2,...,n,...)\}$ underlying the data $\mathbf{Z}_0{:=}(\mathbf{z}_1,...,\mathbf{z}_n)$, and is referred to as *statistical adequacy*. The latter underwrites the statistical reliability of inference by ensuring that the *actual error probabilities* approximate closely the *nominal* (assumed) ones; applying a .05 significance level test, when the actual type I error is closer to .9 will give rise to unreliable evidence. The surest way to create untrustworthy evidence is to apply a .05 α-significance level (nominal) test when the actual type I error probability is closer to .90. It is important to emphasize that invoking Consistent and Asymptotically Normal (CAN) estimators, or using Heteroskedasticity and Autocorrelation Consistent (HAC) Standard Errors often do not address the unreliability of inference problem; Spanos and McGuirk (2001).

2.1 Early macroeconometric modeling

Empirical modeling of macroeconomic time series began in the early 20th century as data-driven modeling of economic fluctuations (business cycle); see Mitchell (1913), Burns and Mitchell (1946). This data-driven modeling was in response to the theory-oriented modeling of business cycles associated with Wicksell, Cassel, Hawtrey, Hayek, Schumpeter, Robertson inter alia; see Haberler (1937). Frisch (1933) blended elements from both approaches to propose a new family of business cycle models in the form of prespecified stochastic difference equations based on a 'propagation' (systematic dynamics) and an 'impulse' (shock) component. This approach inspired Tinbergen to use statistical procedures like least-squares, to estimate the first dynamic macro-econometric model of the Dutch economy in 1936. He extended his empirical framework to compare the empirical validity of the various 'business cycle' theories in Haberler (1937), and estimated a more elaborate macro-model for the USA economy in Tinbergen (1939). Keynes (1939) severely criticized Tinbergen's statistical procedures by raising several foundational problems and issues associated with the use and abuse of linear regression (Hendry and Morgan, 1995), including:

[i] the need to account for all the relevant contributing factors at the outset,

[ii] the conflict between observational data and *ceteris paribus* clauses,

[iii] the spatial and temporal heterogeneity of economic phenomena,

[iv] the validity of the assumed functional forms of economic relations,

[v] the ad hoc specification of the lags and trends in economic relations, and

[vi] the limited applicability of statistical techniques; inappropriate when used with data which cannot be viewed as 'random samples' from a particular population.

In a path breaking paper Haavelmo (1943) proposed a statistical technique to account for the presence of simultaneity bias, calling into question:

[vii] the estimation of a system of interdependent equations using least-squares.

Haavelmo (1944) proposed a most insightful discussion of numerous methodological issues pertaining to empirical modeling in economics, including:

[viii] the need to bridge the gap between the variables envisaged by economic theory and what the available data measure; Spanos (2013).

Unfortunately for econometrics the issues [i]-[viii], with the exception of [vii], have been largely neglected by the subsequent literature.

Despite the disagreements concerning the primary sources of the untrustworthiness of evidence problem, we question (Spanos, 2006a) views like the unreliability of inference problem is not very serious, or the untrustworthiness is primarily due to:

[a] the inherent unpredictability of economic phenomena,

[b] the inevitable fallibility of models; all models are wrong, but some are useful,

[c] the unavoidable price one has to pay for policy-oriented modeling.

By the mid 1950s the macro-econometric models inspired by the Cowles Commission literature (Klein and Goldberger, 1955), showed only marginal improvements on the Tinbergen (1939) model. Moreover, the problem of model validation was confined mainly to securing a high goodness-of-fit (e.g. R^2) and 'error-autocorrelation correction'; see Johnston (1963). Subsequent attempts to ameliorate the adequacy of such macro-econometric models focused primarily on enhancing their scope and 'realisticness' by increasing the number of equations from 15 in the Klein-Goldberger model to several hundred equations of the Brookings and then the Wharton quarterly model of the USA; see Fromm and Klein (1975), McCarthy (1972). Ironically, as these models kept increasing in size the simultaneity and error-autocorrelation corrections were dropped on pragmatic grounds. The ultimate demise of the empirical macro-models of the 1970s and 1980s was primarily due to their bad forecasting performance. When these models were compared with data-driven ARIMA(p,d,q) models, on forecasting grounds, they were found wanting; see Cooper (1972).

2.2 The Pre-Eminence of Theory (PET) perspective

Since Ricardo (1817), theory has generally held the pre-eminent role in economics with data being given the subordinate role of: 'quantifying theories' presumed to be true. Cairnes (1888) articulated an extreme version of the Pre-Eminence of Theory (PET) perspective arguing that data is irrelevant for appraising the 'truth' of economic theories; Spanos (2010b). Robbins (1935) expressed the same view, and the current PET perspective is almost as extreme as Cairnes:

"Any model that is well enough articulated to give clear answers to the questions we put to it will necessarily be artificial, abstract, patently 'unreal'." (Lucas, 1980, p. 696)

"The model economy which better fits the data is not the one used. Rather currently established theory dictates which one is used." (Kydland and Prescott, 1991)

From the PET perspective data does not so much test as allow instantiation of theories: econometric methods offer elaborate (but often misleading) ways 'to bring data into line' with an assumed theory; DSGE modeling is the quintessential example of that; see Spanos (2009). Since the theory has little or no chance to be falsified, such instantiations provide no genuine tests of the theory.

Lucas (1976) and Lucas and Sargent (1981) argued for enhancing the reliance on theory by constructing structural models that are founded directly on the interdependence of a few representative rational agent's [e.g. household, firm, government, central bank] intertemporal optimization (e.g. the maximization of life-time utility) that integrates their expectations directly. The parameters of these models reflect primarily the preference of the decision maker as well as technical and institutional constraints. The claim by Lucas was that such structural models will be invariant to policy interventions and thus provide a better basis for prediction and policy evaluations. This call had widespread appeal and led to the Real Business Cycle (RBC) models which, eventually, culminated in the Dynamic Stochastic General Equilibrium (DSGE) models of today; see Canova (1997), DeJong and Dave (2011), Favero (2001).

2.3 The Error Statistical perspective

Error statistics, a refinement/extension of the Fisher-Neyman-Pearson approach to modeling and inference (Mayo & Spanos, 2004, 2006, 2011), proposes a framework to bridge the gap between theory and data using a *sequence of interconnected models* (theory, structural, statistical, empirical; Spanos, 1986, p. 21), as well as addressing the foundational problems [i]-[x]. The main components of error statistics come in the form of three crucial links between theory and evidence.

A. From an abstract theory T to testable hypotheses h: fashioning an abstract and idealized theory T, initially into a theory model $\mathcal{M}_{\boldsymbol{\psi}}(\mathbf{z};\boldsymbol{\xi})$ which might often include latent variables $\boldsymbol{\xi}$ and then into a structural (substantive) model $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$ that is estimable in light of data $\mathbf{z}_0$. The testable substantive hypotheses of interest h are framed in the context of $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$.

B. From raw data to reliable evidence: $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$ is parametrically nested within a statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$, via $\mathbf{G}(\boldsymbol{\varphi},\boldsymbol{\theta})=\mathbf{0}$, by viewing $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ as a parameterization of the stochastic process $\{\mathbf{Z}_t,\ t\in\mathbb{N}\}$. Reliable 'evidence' $\mathbf{e}$ takes the form of an *adequate statistical model* $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ pertinent for appraising h. $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ is statistically adequate when its probabilistic assumptions are valid for data $\mathbf{Z}_0$.

A statistically adequate model ensures that the actual error probabilities approximate well the *assumed* (nominal) ones.

C. Confronting substantive hypotheses with reliable evidence: probing the substantive hypotheses of interest h only in the context of a statistically adequate model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$. $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ can be used as a reliable benchmark to test the empirical validity of $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$ via $\mathbf{G}(\boldsymbol{\varphi},\boldsymbol{\theta})=\mathbf{0}$, as well as to probe its *substantive adequacy*, i.e. whether $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$ sheds adequate light on (describe, explain) the phenomenon of interest.

3 Statistical vs. Substantive premises of inference

The PET perspective has encouraged modelers to estimate the structural model $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$ directly, ignoring the fact that the statistical premises are specified by an implicit statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ whose invalidity vis-a-vis data $\mathbf{Z}_0$ will *undermine* all inferences based on $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$. This is illustrated in the next subsection.

3.1 Revisiting the Simultaneous Equations Model (SEM)

Consider a generic *Structural Form* (SF) of the *Simultaneous Equations Model (SEM)*:

$$\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z}): \quad \boldsymbol{\Gamma}^{\top}\mathbf{y}_t + \boldsymbol{\Delta}^{\top}\mathbf{x}_t = \boldsymbol{\varepsilon}_t,\ \boldsymbol{\varepsilon}_t \backsim \mathsf{N}(\mathbf{0}, \boldsymbol{\Omega}), \tag{1}$$

with the unknown elements of $\boldsymbol{\Gamma}$, $\boldsymbol{\Delta}$ and $\boldsymbol{\Omega}$ defining the *structural parameters* $\boldsymbol{\varphi}$ of interest. Related to this, is a *Reduced Form* (RF):

$$\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z}): \quad \mathbf{y}_t = \mathbf{B}^{\top}\mathbf{x}_t + \mathbf{u}_t,\ \mathbf{u}_t \backsim \mathsf{N}(\mathbf{0}, \mathbf{V}). \tag{2}$$

where $\mathbf{u}_t = \left(\boldsymbol{\Gamma}^{\top}\right)^{-1}\boldsymbol{\varepsilon}_t$ and $\mathbf{B} = \left(\boldsymbol{\Gamma}^{\top}\right)^{-1}\boldsymbol{\Delta}^{\top}$. The RF is used primarily, if not exclusively, to discuss the *identification* of $\boldsymbol{\varphi}$, via the restrictions:

$$\mathbf{B}(\boldsymbol{\theta})\boldsymbol{\Gamma}(\boldsymbol{\varphi}) + \boldsymbol{\Delta}(\boldsymbol{\varphi}) = \mathbf{0}, \qquad \boldsymbol{\Omega}(\boldsymbol{\varphi}) = \boldsymbol{\Gamma}^{\top}(\boldsymbol{\varphi})\mathbf{V}(\boldsymbol{\theta})\boldsymbol{\Gamma}(\boldsymbol{\varphi}). \tag{3}$$

More generally, the structural and reduced form parameters are related via:

$$\mathbf{G}(\boldsymbol{\varphi}, \boldsymbol{\theta}) = \mathbf{0}. \tag{4}$$

The parameters $\boldsymbol{\varphi} := (\boldsymbol{\Gamma}, \boldsymbol{\Delta})$ are *identified*, if (4) can be solved uniquely for $\boldsymbol{\varphi}$.

Haavelmo (1943) is often credited with pointing out that when least-squares is applied to the structural model (1) yields *inconsistent estimators.* To address the problem he proposed the use of the method of Maximum Likelihood (ML), based on the joint distribution $D(\mathbf{Z}_1, \mathbf{Z}_2, ..\mathbf{Z}_n; \boldsymbol{\theta})$ where $\mathbf{Z}_t := (\mathbf{x}_t, \mathbf{y}_t)$, yielding *consistent* and *parameterization-invariant* estimators of $\boldsymbol{\varphi}$, i.e. the MLE of $\boldsymbol{\varphi}$, $\widehat{\boldsymbol{\theta}}_{MLE}$, can be derived from $\widehat{\boldsymbol{\theta}}_{MLE}$ via: $\mathbf{G}(\widehat{\boldsymbol{\varphi}}_{MLE}, \widehat{\boldsymbol{\theta}}_{MLE}) = \mathbf{0}$, i.e. $\widehat{\boldsymbol{\varphi}}_{MLE}$ is the unique solution of (4), which, under certain probabilistic assumptions, is a *Consistent* and *Asymptotically Normal* (CAN) estimator of $\boldsymbol{\varphi}$.

What is often neglected in traditional econometrics is that the reduced form (2) is the implicit statistical model underwriting the reliability of any inferences based on the estimated $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$. The statistical adequacy of $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ establishes a sound link between $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$ and data $\mathbf{Z}_0$. Hence, inferences based on $\widehat{\boldsymbol{\varphi}}_{MLE}$ can be unreliable if either (a) the reduced form (2) is statistically misspecified, or (b) the overidentifying restrictions in (4) are invalid; Spanos (1990).

$\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ can be validated using thorough *Mis-Specification (M-S) testing* to assess the probabilistic assumptions imposed on the observable process $\{(\mathbf{y}_t|\mathbf{X}_t = \mathbf{x}_t),\ t \in \mathbb{N}\}$. Table 1 specifies (2) in terms of a complete set of testable probabilistic assumptions.

A structural model $\boldsymbol{\Gamma}(\boldsymbol{\varphi})^{\top}\mathbf{y}_t + \boldsymbol{\Delta}(\boldsymbol{\varphi})^{\top}\mathbf{x}_t = \boldsymbol{\varepsilon}_t$, is said to be *empirically valid* when:

(a) the implicit statistical model $\mathbf{y}_t = \mathbf{B}^{\top}(\boldsymbol{\theta})\mathbf{x}_t + \mathbf{u}_t$, is statistically adequate and

(b) the *overidentifying restrictions*: $\mathbf{G}(\boldsymbol{\varphi}, \boldsymbol{\theta}) = \mathbf{0}$ are *data-acceptable*; Spanos (1990).

Table 1 - The Multivariate Linear Regression Model

	Statistical GM:	$\mathbf{y}_t = \boldsymbol{\beta}_0 + \mathbf{B}_1^{\top}\mathbf{x}_t + \mathbf{u}_t, \quad t \in \mathbb{N}.$
[1]	**Normality:**	$D(\mathbf{y}_t|\mathbf{x}_t; \boldsymbol{\theta})$ is Normal
[2]	**Linearity:**	$E(\mathbf{y}_t|\mathbf{X}_t = \mathbf{x}_t) = \boldsymbol{\beta}_0 + \mathbf{B}_1^{\top}\mathbf{x}_t$, linear in $\mathbf{x}_t$,
[3]	**Homoskedasticity:**	$Cov(\mathbf{y}_t|\mathbf{X}_t = \mathbf{x}_t) = \boldsymbol{\Sigma}$, free of $\mathbf{x}_t$,
[4]	**Independence:**	$\{(\mathbf{y}_t|\mathbf{X}_t = \mathbf{x}_t),\ t \in \mathbb{N}\}$ - independ. process.
[5]	**t-invariance:**	$\boldsymbol{\theta} := (\boldsymbol{\beta}_0, \mathbf{B}_1, \mathbf{V})$ are constant for all $t \in \mathbb{N}$.
	$\boldsymbol{\beta}_0 = \boldsymbol{\mu}_1 - \mathbf{B}_1^{\top}\boldsymbol{\mu}_2,\ \mathbf{B}_1 = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21},\ \mathbf{V} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{21}^{\top}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$	

It is important to emphasize that the test in (b) is likely to be unreliable unless one has secured (a). Under (a)-(b) the estimated *empirical model:*

$$\mathcal{M}_{\widehat{\varphi}}(\mathbf{z})\colon \mathbf{\Gamma}(\widehat{\varphi})^{\top}\mathbf{y}_t + \mathbf{\Delta}(\widehat{\varphi})^{\top}\mathbf{x}_t = \widehat{\varepsilon}_t,$$

enjoys both (i) statistical adequacy and (ii) theoretical meaningfulness, and can be used as the basis of reliable inference for prediction and policy simulations.

3.2 The PET's improper implementation of the SEM

The PET implementation of the SEM ignored the statistical adequacy of the reduced form in (a) and explained away the rejection of the overidentifying restrictions in (b) as the inevitable price one has to pay for policy oriented macro-models; see Lucas (1980). Estimating the structural model $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$ directly, and ignoring the adequacy of $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$, gave rise to an improper implementation of the SEM, and as a result the estimated macro-models of the 1970s were both statistically and substantively inadequate. A misspecified $\mathcal{M}_{\widehat{\varphi}}(\mathbf{z})$ will invariably give rise to unreliable inferences and untrustworthy evidence, including non-constant parameter estimates and poor predictive performance.

In retrospect, the poor forecasting performance of these models can be attributed to a number of different factors, the most important of which is that empirical models that do not account for the statistical regularities in the data (statistically misspecified). Such models are likely to give rise to untrustworthy empirical evidence and poor predictive performance; see Granger and Newbold (1986), p. 280. However, this particular source of unreliability was ignored by the new classical macroeconomics of the 1980s, and instead blamed the poor forecasting performance solely on the their ad hoc specification and their lack of proper theoretical microfoundations; see DeJong and Dave (2011).

The adherents of the PET perspective, however, offered a very different explanation for this predictive failure, and instead blamed the substantive inadequacy of these models. This led to the Real Business Cycle (RBC) and DSGE) models:

"... the use of calibration exercises as a means for facilitating the empirical implementation of DSGE models arose in the aftermath of the demise of system of equations analyses." (DeJong and Dave, 2011, p. 257)

3.3 Where do statistical models come from?

Traditionally the probabilistic structure of the statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ is specified indirectly by attaching errors (shocks) to the behavioral equations comprising the structural model $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$. This implicit statistical model is often statistically misspecified because the probabilistic structure imposed on the observable process $\{\mathbf{Z}_t,\ t\in\mathbb{N}\}$ underlying data $\mathbf{Z}_0$, often ignores crucial statistical information contained in the data.

To make any progress one needs to disentangle the *statistical* $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ from the *substantive premises* $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$, without compromising the integrity of either source of information; Spanos (2006b). This can be achieved by viewing $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ as a *parameterization* of the process $\{\mathbf{Z}_t,\ t\in\mathbb{N}\}$ underlying $\mathbf{Z}_0$.

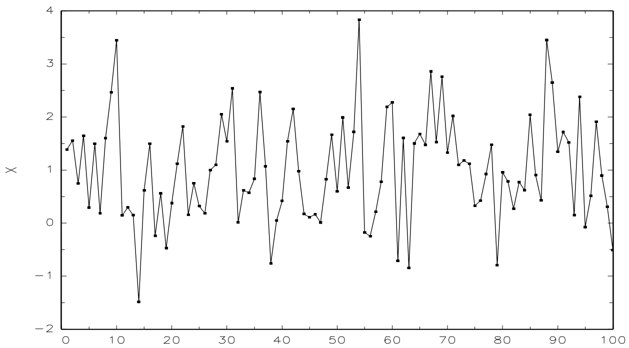


Fig. 3

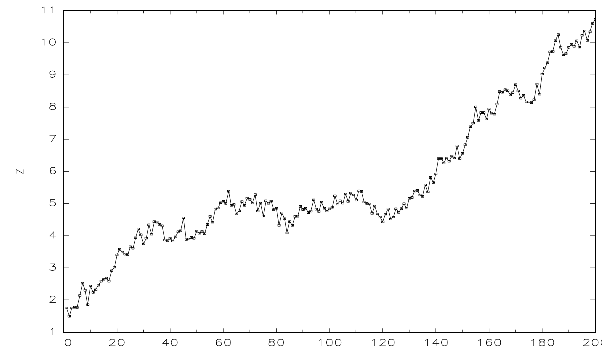


Fig. 4

The construction of $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ begins with a given data $\mathbf{Z}_0$, irrespective of the theory that led to the choice of $\mathbf{Z}_0$. Once selected, data $\mathbf{Z}_0$ take on 'a life of its own' as a particular realization of a generic process $\{\mathbf{Z}_t,\ t{\in}\mathbb{N}\}$. The link between data $\mathbf{Z}_0$ and the process $\{\mathbf{Z}_t,\ t{\in}\mathbb{N}\}$ is provided by a pertinent answer to the key question: 'what probabilistic structure, when imposed on the process $\{\mathbf{Z}_t,\ t{\in}\mathbb{N}\}$, would render data $\mathbf{Z}_0$ a truly *typical realization* thereof?' A 'typical realization' of NIID process $\{Z_t,\ t{\in}\mathbb{N}\}$ looks like fig. 3, not 4! Fig. 4 is a typical realization of a Normal, Markov, mean-trending process.

Step 1 - typicality. The 'truly typical realization' answer provides the relevant probabilistic structure for $\{\mathbf{Z}_t,\ t{\in}\mathbb{N}\}$; an answer that can be empirically assessed using thorough Mis-Specification (M-S) testing.

Step 2 - parameterization. The relevant statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ is specified by choosing a particular parameterization $\boldsymbol{\theta}{\in}\Theta$ for $\{\mathbf{Z}_t,\ t{\in}\mathbb{N}\}$, with a view to nest parametrically the structural model $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$, e.g. $\mathbf{G}(\boldsymbol{\theta},\boldsymbol{\varphi}){=}\mathbf{0}$, $\boldsymbol{\varphi}{\in}\Phi$.

Example. For the data in fig. 2, a particular parameterization $\boldsymbol{\theta}{:=}(\alpha_0,\delta_1,\delta_2,\alpha_1,\sigma^2)$, of $\{Z_t,\ t{\in}\mathbb{N}\}$ gives rise to *the Normal AR(1) with a trend:*

$$\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z}){:}\ (Z_t|Z_{t-1}){\backsim}\mathsf{N}(\alpha_0+\delta_1 t+\delta_2 t^2+\alpha_1 Z_{t-1},\sigma_0^2),\quad t{\in}\mathbb{N}.$$

More generally, data $\mathbf{Z}_0$ is viewed as a realization of a generic (vector) stochastic process $\{\mathbf{Z}_t,\ t{\in}\mathbb{N}\}$, regardless of what the variables $\mathbf{Z}_t$ measure substantively. This disentangling enables one to delineate between the two questions (Spanos, 2006c):

Statistical adequacy: does $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ account for the chance regularities in $\mathbf{Z}_0$?

Substantive adequacy: does the model $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$ shed adequate light (describe, explain, predict) on the phenomenon of interest?

Establishing the statistical adequacy of $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ first, enables one to ensure the reliability of any inference pertaining to the substantive questions of interest, including the *validity* of the restrictions $\mathbf{G}(\boldsymbol{\theta},\boldsymbol{\varphi}){=}\mathbf{0}$, $\boldsymbol{\varphi}{\in}\Phi$, $\boldsymbol{\theta}{\in}\boldsymbol{\Theta}$.

4 Revisiting DSGE modeling

Dynamic Stochastic General Equilibrium (DSGE) models are currently dominating both empirical modeling in macroeconomics as well as policy evaluation; see Canova (2007). The DSGE models start from life time optimization problem faced by consumers and firms. The first order conditions of the optimization problem are highly non linear in level variables. These conditions are linearized around constant steady state using first order Taylor approximation, which is a local approximation and can be misleading. Using second order approximation raises different problems because when linearity is lost the Kalman filter cannot be used; see Heer and Maussner (2009), DeJong and Dave (2011). After linearization, the model is in terms of log difference, which is thought to be substantively more meaningful.

A typical small DSGE model $\mathcal{M}_{\boldsymbol{\psi}}(\mathbf{z};\boldsymbol{\xi};\boldsymbol{\epsilon}_t)$ based on Ireland (2004, 2011), after linearization is expressed in terms of three types of variables (table 2):

(i) observables (Y_t, P_t, r_t), Y_t-production, P_t-price level, and r_t- interest rate,
(ii) latent variables $\boldsymbol{\xi}_t{=}(Q_t, g_t)$, Q_t-efficient output, $g_t{=}(Y_t/Q_t)$-output gap,
(iii) latent shocks: $\boldsymbol{\epsilon}_t{=}(a_t, \theta_t, H_t)$, a_t-preference, θ_t-demand, H_t-technology.

The estimable form of $\mathcal{M}_{\boldsymbol{\psi}}(\mathbf{z};\boldsymbol{\xi};\boldsymbol{\epsilon}_t)$, the *structural* DSGE model $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$, is derived by solving a system of linear expectational difference equations and eliminating certain variables. $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$ is specified in terms of the observables $\mathbf{Z}_t{:=}(\widehat{y}_t, \widehat{p}_t, \widehat{r}_t)$:

$$\widehat{y}_t{=}\ln(Y_t/Y_{t-1}){-}\ln(\overline{y}),\ \widehat{p}_t{=}\ln(P_t/P_{t-1}){-}\ln(\overline{p}),\ \widehat{r}_t{=}\ln(r_t){-}\ln(\overline{r}).$$

Table 2: Dynamic Stochastic General Equilibrium (DSGE) model

Behavioral equations:

$$\text{(i)}\ \ln(\tfrac{Y_t/Q_t}{\overline{g}}){=}\alpha_g\ln(\tfrac{Y_{t-1}/Q_{t-1}}{\overline{g}}){+}(1{-}\alpha_g)E_t\ln(\tfrac{Y_{t+1}/Q_{t+1}}{\overline{g}}){-}\left\{\ln\left(\tfrac{r_t}{\overline{r}}\right)-E_t[\ln(\tfrac{P_{t+1}/P_t}{\overline{p}})]\right\}+$$
$$+(1{-}\omega)(1{-}\rho_a)\ln(a_t),\qquad \text{where } \overline{w}{=}\tfrac{1}{n}\textstyle\sum_{t=1}^{n}(W_t/W_{t-1})$$

$$\text{(ii)}\ \ln(\tfrac{P_t/P_{t-1}}{\overline{p}}){=}\beta\left(\alpha_p\ln(\tfrac{P_{t-1}/P_{t-2}}{\overline{p}})\right)+(1{-}\alpha_p)E_t\ln(\tfrac{P_{t+1}/P_t}{\overline{p}}){+}\psi\ln(\tfrac{Y_t/Q_t}{\overline{g}})\text{-}(\tfrac{1}{\phi})\ln(\tfrac{\theta_t}{\overline{\theta}})$$

$$\text{(iii)}\ \ln(\tfrac{r_t}{\overline{r}})-\ln(\tfrac{r_{t-1}}{\overline{r}}){=}\rho_p\ln(\tfrac{P_t/P_{t-1}}{\overline{p}})+\rho_y\ln(\tfrac{Y_t/Y_{t-1}}{\overline{y}})+\rho_g\ln(\tfrac{Y_t/Q_t}{\overline{g}})+\varepsilon_{rt}$$

$$\text{(iv)}\ \ln(\tfrac{Y_t/Q_t}{\overline{g}}){=}\ln(\tfrac{Y_t/H_t}{\overline{y}_h})\text{-}\omega\ln(a_t)\quad \text{(v)}\ \ln(\tfrac{Y_t/Y_{t-1}}{\overline{y}}){=}\ln(\tfrac{Y_t/H_t}{\overline{y}_h})\text{-}\ln(\tfrac{Y_{t-1}/H_{t-1}}{\overline{y}_h}){+}\ln(\tfrac{H_t/H_{t-1}}{\overline{h}})$$

$$\textbf{Shocks}\begin{cases}\ln a_t{=}\rho_a\ln a_{t-1}{+}\varepsilon_{at},\ \ \ln(\tfrac{\theta_t}{\overline{\theta}}){=}\ln\phi{+}\tfrac{\rho_\theta}{\phi}\ln\tfrac{\theta_{t-1}}{\overline{\theta}}{+}\varepsilon_{\theta t},\ \ \ln(\tfrac{H_t/H_{t-1}}{\overline{h}}){=}\varepsilon_{ht}\\ \boldsymbol{\varepsilon}_t{:=}(\varepsilon_{at},\varepsilon_{\theta t},\varepsilon_{ht},\varepsilon_{rt})\backsim \mathsf{NIID}(\mathbf{0},\boldsymbol{\Lambda}{=}\mathrm{diag}\,(\sigma_\alpha^2,\sigma_\theta^2,\sigma_h^2,\sigma_r^2)),\end{cases}$$

Parameters: $\boldsymbol{\varphi}{:=}\left(\alpha_g,\rho_a,\beta,\psi,\alpha_p,\phi,\rho_p,\rho_y,\rho_g,\rho_\theta,\omega,\psi,\sigma_\alpha^2,\sigma_\theta^2,\sigma_\eta^2,\sigma_r^2\right)$

DSGE models aim to describe the behavior of the economy in an equilibrium steady state stemming from optimal microeconomic decisions associated with several representative agents (households, firms, governments, central banks). It is essentially a *deterministic* theory-models in the form of a system of first order difference equations, but driven by *latent* stochastic (autocorrelated) shocks.

4.1 Calibration: model quantification and validation

The model in Table 5 can be expressed as a system:

$$\mathbf{A}\mathbf{E}_t\mathbf{s}_{t+1}=\mathbf{B}\mathbf{s}_t+\mathbf{C}\mathbf{v}_t,\quad \mathbf{v}_t=\mathbf{P}\mathbf{v}_{t-1}+\boldsymbol{\varepsilon}_t,\ \boldsymbol{\varepsilon}_t\sim\ \mathsf{NIID}(\mathbf{0},\mathbf{V}),$$
$$\mathbf{s}_t{=}\left[\ \widehat{y}_{ht-1}\ \ \widehat{r}_{t-1}\ \ \widehat{p}_{t-1}\ \ \widehat{y}_{t-1}\ \ \widehat{g}_{t-1}\ \ \widehat{p}_t\ \ \widehat{g}_t\ \right]^\top,\quad \mathbf{v}_t{=}\left[\widehat{a}_t\ \ \widehat{e}_t\ \ \widehat{h}_t\ \ \varepsilon_{rt}\right]^\top,$$
$$\boldsymbol{\varepsilon}_t{=}\left[\varepsilon_{at}\ \ \varepsilon_{et}\ \ \varepsilon_{ht}\ \ \varepsilon_{rt}\right]^\top,\qquad \mathbf{V}{=}\mathrm{diag}\left(\sigma_\alpha^2,\sigma_\theta^2,\sigma_\eta^2,\sigma_r^2\right),$$

the explicit form of the structural matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{P}$ are given in the Appendix. The 'solution' of the DSGE model using Klein's (2000) algorithm, for:

$$\mathbf{Z}_t{:=}(\widehat{y}_{t-1}, \widehat{p}_{t-1}, \widehat{r}_{t-1}, \widehat{y}_{ht-1}, \widehat{g}_{t-1}, \widehat{a}_t, \widehat{e}_t, \hat{h}_t, \varepsilon_{rt}) \text{ and } \mathbf{X}_t{:=}(\widehat{y}_t, \widehat{p}_t, \widehat{r}_t),$$

yields the restricted state-space formulation:

$$\begin{gathered}\mathbf{Z}_t{=}\mathbf{A}_1(\boldsymbol{\varphi})\mathbf{Z}_{t-1} + \mathbf{A}_2(\boldsymbol{\varphi})\boldsymbol{\varepsilon}_t,\;\; \mathbf{X}_t{=}\mathbf{H}_1(\boldsymbol{\varphi})\mathbf{Z}_t,\\ \boldsymbol{\Omega}(\varphi){=}\mathbf{A}_2(\boldsymbol{\varphi})E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top)\mathbf{A}_2^\top(\boldsymbol{\varphi}),\;\; E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top){=}\,\mathbf{V},\end{gathered} \tag{5}$$

which provides the basis for calibration; note that $\mathbf{A}_1(\varphi), \mathbf{A}_2(\varphi), \mathbf{H}_1(\varphi)$ are defined by Klein's (2000) algorithm (Canova, 2007).

Step 1. Select 'theoretically meaningful' values of all the structural parameters:

$$\boldsymbol{\varphi}{:=}(\omega, \alpha_g, \rho_a, \beta, \psi, \alpha_\pi, \phi, \rho_\pi, \rho_y, \rho_g, \rho_\theta, \omega{=}(1/\eta), \psi{=}\,[\eta(\theta{-}1)/\phi], \sigma^2_\alpha, \sigma^2_\theta, \sigma^2_\eta)$$

Step 2. Select the sample size, say n and the initial values $\mathbf{x}_0$.

Step 3. Use the values in steps 1-2, together with Normal pseudo-random numbers for $\boldsymbol{\varepsilon}_t$ to simulate (5), N runs of size n.

Step 4. After de-trending using the Hodrick-Prescott (H-P) filter, use the simulated data $\mathbf{Z}_0^s$ to evaluate the moment statistics (mean, variances, covariances) of interest for each run of size n, as well as their empirical distributions for all N.

Step 5. Compare the relevant moments of the simulated data $\mathbf{Z}_0^s$ with those of the actual data $\mathbf{Z}_0$, finessing the original values of $\boldsymbol{\varphi}$ to ensure that these moments are close to each other:

$$\min_{\boldsymbol{\varphi}\in\boldsymbol{\Phi}} \parallel Cov(\mathbf{Z}_0^s; \boldsymbol{\varphi}) - Cov(\mathbf{Z}_0) \parallel$$

as well as the model yields realistic-looking data; simulated mimic actual data.

Calibration: $\overline{z}{=}\overline{y}{=}1.0048$, $\overline{p}{=}1.0086$, $\overline{R}{=}\,\frac{1}{n}\sum_{t=1}^n R_t{=}\frac{\overline{p}}{\overline{z}\beta}{\rightarrow}\,\beta{=}.99$, $\psi{=}.1$.

4.2 Confronting the DSGE model with data

Data: US quarterly time series for the period 1948-2010 (n=252): Y_t - per capita real GDP, P_t - GDP deflator, r_t - gross interest rate on 90 days Treasury bill. The validation of the DSGE structural model $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$ takes three steps. Step 1. Unveil the statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ implicit in the DSGE model $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$. Step 2. Secure its statistical adequacy of $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ using M-S testing and respecification. Step 3. Test the overidentifying restrictions in the context of a statistically adequate model.

The implicit statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ behind $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$ is a Normal, VAR(2) model (table 3) in terms of the observables: $\widehat{y}_t{=}\ln(Y_t/Y_{t-1})$, $\widehat{p}_t{=}\ln(P_t/P_{t-1})$, $\widehat{r}_t{=}\ln(r_t)$.

For the link between the structural and the statistical model in tables 2-3, see appendix.

Table 3: Normal VAR(2) model

	Statistical GM:	$\mathbf{Z}_t = \mathbf{a}_0 + \mathbf{A}_1^\top\mathbf{Z}_{t-1} + \mathbf{A}_2^\top\mathbf{Z}_{t-2} + \mathbf{u}_t,\ t{\in}\mathbb{N},$
[1]	Normality:	$D(\mathbf{Z}_t, \mathbf{Z}_{t-1}, ...\mathbf{Z}_1; \boldsymbol{\theta})$, is Normal
[2]	Linearity:	$E(\mathbf{Z}_t\vert\sigma(\mathbf{Z}_{t-1}^0)) = \mathbf{a}_0 + \mathbf{A}_1^\top\mathbf{Z}_{t-1} + \mathbf{A}_2^\top\mathbf{Z}_{t-2},$
[3]	Homosked.:	$Var(\mathbf{Z}_t\vert\sigma(\mathbf{Z}_{t-1}^0)){=}\mathbf{V}$ is free of $\mathbf{Z}_{t-1}^0{:=}(\mathbf{Z}_{t-1}, ...\mathbf{Z}_1)$,
[4]	Markov:	$\{\mathbf{Z}_t,\ t{\in}\mathbb{N}\}$ is a Markov(2) process
[5]	t-invariance:	$\boldsymbol{\theta}{:=}(\mathbf{a}_0, \mathbf{A}_1, \mathbf{A}_2, \mathbf{V})$ are t-invariant for all $t{\in}\mathbb{N}$.

The auxiliary regressions to test the assumptions [2]-[5] are written in terms of the standardized residual of the growth rate equation ($\widehat{u}_{yt}$) of the VAR(2).

$$\widehat{u}_{yt} = a_0 + a_1\widehat{u}_{yt-1} + a_2\widehat{u}_{yt-2} + v_t \tag{6}$$

$$\widehat{u}_{yt} = b_0 + b_1\widehat{y}_t + b_2\widehat{y}_t^2 + b_3t + b_4t^2 + v_t \tag{7}$$

$$\widehat{u}_{yt}^2 = c_0 + c_1\widehat{y}_t + c_2\widehat{y}_t^2 + c_3\widehat{y}_{t-1}^2 + c_4\widehat{y}_{t-2}^2 + c_5t + c_6t^2 + v_t \tag{8}$$

The form of the auxiliary regressions being used for joint M-S testing depends on a number of different factors and the robustness of the results is evaluated using several alternative forms.

Table 4: Model assumption	**Null Hypotheses**
Linearity F(242,1)	$H_0 : b_2 = 0$
t-invariance F(242,2)	$H_0 : b_3 = b_4 = 0$
Independence F(242,2)	$H_0 : a_1 = a_2 = 0$
Homoskedasticity F(238,2)	$H_0 : c_1 = c_2 = 0$
2nd order Independence F(238,2)	$H_0 : c_3 = c_4 = 0$
2nd order t-invariance F(238,2)	$H_0 : c_5 = c_6 = 0$

Table 5: M-S testing results: Normal VAR(2) model

Model assumption	y_t	p_t	r_t
Normality	.982[.008]*	.901[.000]*	.791[.000]*
Linearity	1.44[.232]	0.607[.437]	.070[.792]
Homoskedasticity	5.299[.006]*	37.285[.000]*	3.401[.035]*
Markov(2)	0.348[.706]	3.488[.032]*	11.624[.000]*
t-invariance	12.008[.000]*	50.542[.000]*	2.593[.077]

The M-S results reported in table 5 (p-values in square brackets) indicate that the estimated VAR(2) model is *seriously misspecified*; the validity of all assumptions but [2] are called into question. Hence, no reliable inferences can be drawn on the basis the estimated VAR(2), including testing the validity of the DSGE restrictions! The next step is to respecify this model with a view to account for the statistical information not accounted for by the VAR(2) model.

Table 6: Student's t VAR(3) model

	Statistical GM:	$\mathbf{Z}_t = \mathbf{a}_0(t) + \mathbf{A}_1^\top\mathbf{Z}_{t-1} + \mathbf{A}_2^\top\mathbf{Z}_{t-2} + \mathbf{A}_3^\top\mathbf{Z}_{t-3} + \mathbf{u}_t,\ t\in\mathbb{N},$
[1]	Normality:	$D(\mathbf{Z}_t, \mathbf{Z}_{t-1}, ...\mathbf{Z}_1; \nu; \boldsymbol{\theta})$, is Student's t with ν d.f.
[2]	Linearity:	$E(\mathbf{Z}_t\vert\sigma(\mathbf{Z}_{t-1}^0))=\mathbf{a}_0(t) + \mathbf{A}_1^\top\mathbf{Z}_{t-1} + \mathbf{A}_2^\top\mathbf{Z}_{t-2} + \mathbf{A}_3^\top\mathbf{Z}_{t-3},$
[3]	Homosked.:	$Var(\mathbf{Z}_t\vert\sigma(\mathbf{Z}_{t-1}^0))=q(\mathbf{Z}_{t-1}^0)$ is free of $\mathbf{Z}_{t-1}^0{:=}(\mathbf{Z}_{t-1}, ...\mathbf{Z}_1),$
		$q(\mathbf{Z}_{t-1})=\left(\frac{\nu}{\nu+m-2}\right)\mathbf{V}[1 + \frac{1}{\nu}\sum_{i=1}^{3}(\mathbf{Z}_{t-i} - \boldsymbol{\mu})\mathbf{Q}_i^{-1}(\mathbf{Z}_{t-i} - \boldsymbol{\mu})]$
[4]	Markov:	$\{\mathbf{Z}_t,\ t\in\mathbb{N}\}$ is a Markov process
[5]	t-invariance:	$\boldsymbol{\theta}{:=}(\mathbf{a}_0, \boldsymbol{\mu}, \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \mathbf{V}, \mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3)$ are constant for $t\in\mathbb{N}$.

Respecification. The non-Normality, Heteroskedasticity and second-order temporal dependence suggest replacing the original Normality with a another distribution from the Elliptically Symmetric family. In light of this diagnosis, the process $\{\mathbf{Z}_t,\ t\in\mathbb{N}\}$ is now assumed to be Student's, Markov and Stationary, giving rise to the Student's VAR(3) [St-VAR(3)] model (table 6). To be fair, the implicit Normal VAR(2) [N-VAR] model allowed for the possibility of MA(1) errors, which can justify the third lag in the St-VAR(3) model.Estimation of the St-VAR(3) model (see Appendix) yielded the results in table 7, which are contrasted to those of the N-VAR(2) in table 8.

	Table 7: Student's t VAR(3)			**Table 8: Normal VAR(2)**		
	y_t	p_t	r_t	y_t	p_t	r_t
δ_0	.493[.000]	-.014[.545]	-.159[.000]	.526[0.000]	.041[.458]	.003[.893]
t	-.159[.000]	.108[.002]	.078[.000]			
t^2	1.060[.000]	-.468[.001]	-.377[.000]			
y_{t-1}	.285[.000]	.022[.178]	.023[.000]	.305[.000]	.038[.185]	.025[.040]
p_{t-1}	.287[.023]♦	.456[.000]	.026[.146]	.193[.191]♦	.549[.000]	.032[.232]
r_{t-1}	$-$.607[.061]♦	.166[.187]	1.359[.000]	.015[.360]♦	.243[.117]	1.126[.000]
y_{t-2}	.110[.027]	.004[.848]	.005[.524]	.096[.137]	.020[.476]	.012[.325]
p_{t-2}	-.238[.168]	.176[.005]	.012[.597]	-.243[.104]	.179[.006]	.035[.203]
r_{t-2}	.273[.671]	-.305[.199]♦	-.680[.000]♦	-.197[.582]	-.114[.459]♦	-.191[.004]♦
y_{t-3}	-.205[.001]	.012[.585]	-.003[.774]			
p_{t-3}	-.446[.021]	.222[.001]	-.011[.672]			
r_{t-3}	.215[.606]	.167[.288]	.265[.000]			

The key difference between the two models comes in the form of the conditional variance which plays a crucial role in rendering the St-VAR(3) model adequate.

Table 9: $\widehat{Var}(y_t|\mathbf{Z}^0_{t-1})$ for $\widetilde{x}_t{:=}(x_t{-}\widehat{E(x_t)})$

$$\begin{aligned}
&\underset{(.000)}{.133} + \underset{(.000)}{.104}\tilde{y}^2_{t-1} + \underset{(.000)}{.131}\tilde{y}_{t-1}\tilde{p}_{t-1} - \underset{(.000)}{.344}\tilde{y}_{t-1}\tilde{r}_{t-1} - \underset{(.000)}{.054}\tilde{y}_{t-1}\tilde{y}_{t-2} - \underset{(.000)}{.099}\tilde{y}_{t-1}\tilde{p}_{t-2}\\
&+\underset{(.000)}{.562}\tilde{y}_{t-1}\tilde{r}_{t-2} - \underset{(.268)}{.011}\tilde{y}_{t-1}\tilde{y}_{t-3} + \underset{(.036)}{.056}\tilde{y}_{t-1}\tilde{p}_{t-3} - \underset{(.001)}{.207}\tilde{y}_{t-1}\tilde{r}_{t-3} + \underset{(.000)}{.773}\tilde{p}^2_{t-1}\text{-}\underset{(.000)}{.866}\tilde{p}_{t-1}\tilde{r}_{t-1}\\
&\text{-}\underset{(.082)}{.041}\tilde{p}_{t-1}\tilde{y}_{t-2}\text{-}\underset{(.000)}{.799}\tilde{p}_{t-1}\tilde{p}_{t-2} + \underset{(.000)}{1.074}\tilde{p}_{t-1}\tilde{r}_{t-2}\text{-}\underset{(.459)}{.019}\tilde{p}_{t-1}\tilde{y}_{t-3}\text{-}\underset{(.000)}{.377}\tilde{p}_{t-1}\tilde{p}_{t-3}\text{-}\underset{(.085)}{.303}\tilde{p}_{t-1}\tilde{r}_{t-3}\\
&+\underset{(.000)}{4.885}\tilde{r}^2_{t-1} - \underset{(.007)}{.160}\tilde{y}_{t-1}\tilde{y}_{t-2} - \underset{(.286)}{.179}\tilde{y}_{t-1}\tilde{p}_{t-2} - \underset{(.000)}{12.499}\tilde{y}_{t-1}\tilde{r}_{t-2} - \underset{(.222)}{.074}\tilde{y}_{t-1}\tilde{y}_{t-3}+\\
&\underset{(.373)}{.155}\tilde{y}_{t-1}\tilde{p}_{t-3} - \underset{(.000)}{3.428}\tilde{y}_{t-1}\tilde{r}_{t-3} + \underset{(.000)}{.111}\tilde{y}^2_{t-2} + \underset{(.000)}{.176}\tilde{y}_{t-2}\tilde{p}_{t-2} - \underset{(.257)}{.111}\tilde{y}_{t-2}\tilde{r}_{t-2} - \underset{(.000)}{.047}\tilde{y}_{t-2}\tilde{y}_{t-3}-\\
&-\underset{(.003)}{.073}\tilde{y}_{t-2}\tilde{p}_{t-3} - \underset{(.000)}{.278}\tilde{y}_{t-2}\tilde{r}_{t-3} + \underset{(.000)}{.934}\tilde{p}^2_{t-2} - \underset{(.002)}{.929}\tilde{p}_{t-2}\tilde{r}_{t-2} - \underset{(.184)}{.033}\tilde{p}_{t-2}\tilde{y}_{t-3} - \underset{(.000)}{.794}\tilde{p}_{t-2}\tilde{p}_{t-3}\\
&+\underset{(.000)}{.665}\tilde{p}_{t-2}\tilde{r}_{t-3} + \underset{(.000)}{12.288}\tilde{r}^2_{t-2} - \underset{(.000)}{.359}\tilde{r}_{t-2}\tilde{y}_{t-3} - \underset{(.816)}{.064}\tilde{r}_{t-2}\tilde{p}_{t-3} - \underset{(.000)}{12.454}\tilde{r}_{t-2}\tilde{r}_{t-3} + \underset{(.000)}{.100}\tilde{y}^2_{t-3}+\\
&+\underset{(.000)}{.130}\tilde{y}_{t-3}\tilde{p}_{t-3} + \underset{(.000)}{.264}\tilde{y}_{t-3}\tilde{r}_{t-3} + \underset{(.000)}{.783}\tilde{p}^2_{t-3} - \underset{(.103)}{.254}\tilde{p}_{t-3}\tilde{r}_{t-3} + \underset{(.000)}{4.827}\tilde{y}^2_{t-3}
\end{aligned}$$

The estimated $\widehat{Var}(y_t|\mathbf{Z}^0_{t-1})$ (table 9) shows numerous highly significant terms and can be used to explain the manifest differences between the Student's t and Normal VAR estimates in tables 7 and 8. The differences are clearly due to the conditional variance $Var(\mathbf{y}_t|\mathbf{Z}_{t-1}^0)$ being heteroskedastic and the parameters of the autoregressive and autoskedastic functions being interrelated via the parameters of the joint distribution. In addition, the estimated $\widehat{Var}(y_t|\mathbf{Z}_{t-1}^0)$ brings out the potential unreliability of any impulse response and variance decomposition analysis based on assuming a constant conditional variance.

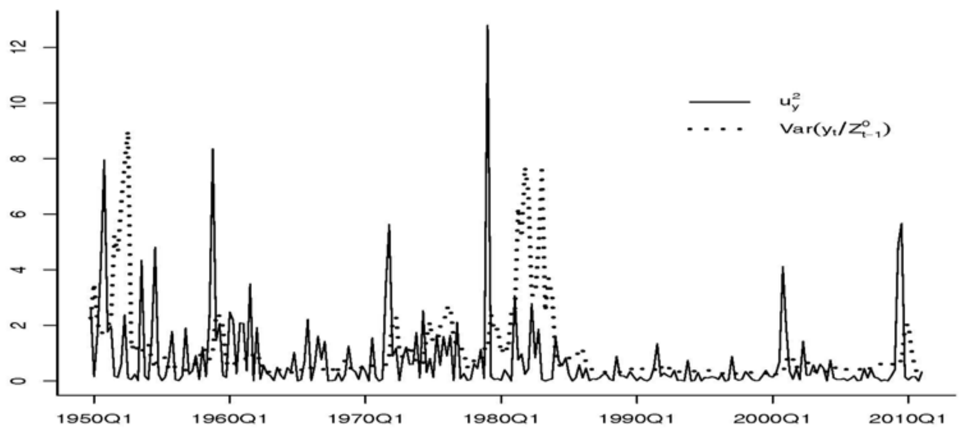


Fig. 5: N-VAR residuals $(\widehat{u}_{1t}^2)$ vs. $\widehat{Var}(y_t|\mathbf{Z}_{t-1}^0)$

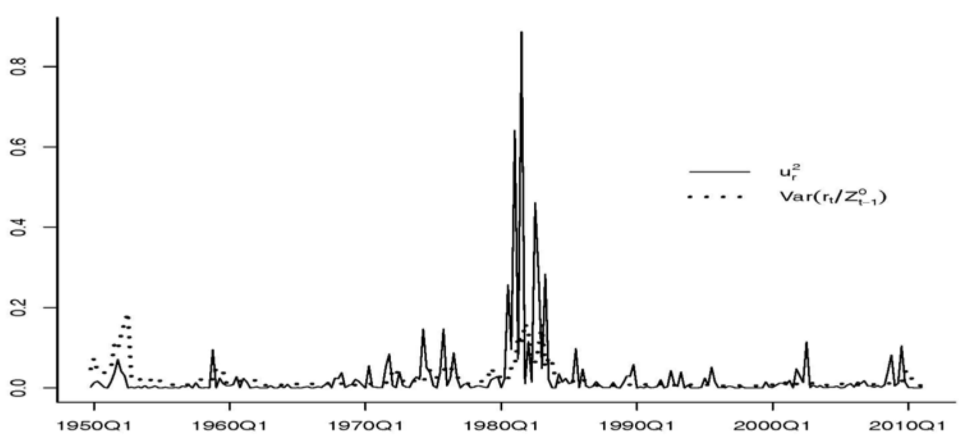


Fig. 6: N-VAR residuals $(\widehat{u}_{2t}^2)$ vs. $\widehat{Var}(r_t|\mathbf{Z}_{t-1}^0)$

The inappropriateness of a constant conditional variances is illustrated in figures 5-6, contrasting the N-VAR(2) squared residuals and $\widehat{Var}(y_t|\mathbf{Z}_{t-1}^0)$ of the St-VAR(3) (table 9); NOTE that all three conditional variances are scaled versions of each other.

To take into account the heteroskedastic conditional variance-covariance, one needs to reconsider the notion of what constitutes the relevant residuals for M-S testing purposes which should be defined in terms of the standardized residuals:

$$\widehat{\mathbf{u}}_t = \left(\begin{array}{ccc} \widehat{u}_{yt} & \widehat{u}_{pt} & \widehat{u}_{rt} \end{array} \right)^\top = \mathbf{L}_t^{-1}(\mathbf{Z}_t - \widehat{\mathbf{Z}}_t)$$

where $\mathbf{L}_t\mathbf{L}_t^\top = \widehat{Var}(\mathbf{Z}_t|\boldsymbol{\sigma}(\mathbf{Z}_{t-1}^0))$. Here, $\mathbf{L}_t$ is changing with t and $\mathbf{Z}_{t-1}^0$ as opposed to the constant conditional variance-covariance in the case of the stationary Normal VAR model. An indicative set of auxiliary regressions based on these residuals is:

$$\widehat{u}_{yt} = a_0 + a_1\widehat{u}_{yt-1} + a_2\widehat{u}_{yt-2} + b_1\widehat{y}_t + b_2\widehat{\widehat{y}}_t^2 + b_3t^3 + b_4t^4 + v_{1t}$$
$$\widehat{u}_{yt}^2 = c_0 + c_1\widehat{\sigma}_{yt} + c_2\widehat{\widehat{\sigma}}_t^2 + c_4\widehat{\sigma}_{yt-1} + c_5\widehat{\sigma}_{yt-2} + d_1t^5 + d_2t^6 + v_{2t}$$
$$\widehat{\sigma}_{yt}^2 = \widehat{Var}(y_t|\sigma(\mathbf{Z}_{t-1}^0)),\ \widehat{\widehat{\sigma}}_{yt}^2 = \left(\tfrac{\nu}{\nu+3k-2}\right)\widehat{\omega}_y[1+\tfrac{1}{\nu}(\mathbf{Z}_{t-1}^0-\widehat{\boldsymbol{\mu}}_0)\widehat{\mathbf{Q}}^{-1}(\mathbf{Z}_{t-1}^0-\widehat{\boldsymbol{\mu}}_0)])$$

$$\left(\begin{array}{ccc} \widehat{y}_t & \widehat{p}_t & \widehat{r}_t \end{array} \right)^\top = \widehat{\boldsymbol{\delta}}_0 + \widehat{\boldsymbol{\delta}}_1 t + \widehat{\boldsymbol{\delta}}_2 t^2 + \widehat{\mathbf{A}}_1^\top\mathbf{Z}_{t-1} + \widehat{\mathbf{A}}_2^\top\mathbf{Z}_{t-2} + \widehat{\mathbf{A}}_3^\top\mathbf{Z}_{t-3}$$
$$\left(\begin{array}{ccc} \widehat{\widehat{y}}_t & \widehat{\widehat{p}}_t & \widehat{\widehat{r}}_t \end{array} \right)^\top = \widehat{\boldsymbol{\delta}}_0 + \widehat{\mathbf{A}}_1^\top\mathbf{Z}_{t-1} + \widehat{\mathbf{A}}_2^\top\mathbf{Z}_{t-2} + \widehat{\mathbf{A}}_3^\top\mathbf{Z}_{t-3}$$

$\widehat{y}_t$ - fitted values, represents the linear combination of the terms in the conditional mean from the null model. On the other hand, $\widehat{\widehat{y}}_t$ represents the fitted values minus the trend terms so that $\widehat{\widehat{y}}_t^2$ represents the pure departure from the linearity assumption. Similarly, $\widehat{\sigma}_t$ represents the linear combination of the quadratic terms on the right hand side of the conditional variance $\widehat{var}(y_t|\sigma(\mathbf{Z}_{t-1}^0))$. $\widehat{\widehat{\sigma}}_{yt}$ represents the estimated $\widehat{var}(y_t|\sigma(\mathbf{Z}_{t-1}^0))$ minus the trend components. In other words, $\widehat{\widehat{\sigma}}_{yt}$ represents the pure heteroskedastic (i.e. the terms depending only on $\mathbf{Z}_{t-1}^0$) term of the conditional variance so that $\widehat{\widehat{\sigma}}_{yt}^2$ represents pure departure from the assumption of quadratic heteroskedasticity. This strategy allows us to test the t-invariance assumption separately from the assumptions of heteroskedasticity and second order dependence. The hypotheses being tested are directly analogous to those in table 4 above.

Thorough M-S testing of the estimated Student's VAR(3) model indicates no departures from its assumptions; see table 10. The statistical adequacy is reflected in the constancy of the variation around the mean exhibited by the St-VAR(3) residuals in fig. 8, in contrast to the N-VAR(2) residuals in fig. 7.

Table 10: M-S testing results: Student's t, VAR(3) model

Model assumption	y_t	p_t	r_t
Student's t	2.061[.357]	3.200[.202]	1.351[.509]
Linearity	1.378[.254]	0.076[.927]	1.465[.233]
Heteroskedasticity	1.508[.221]	0.890[.347]	4.222[.051]
Independence	0.335[.716]	0.417[.660]	2.693[.070]
t-invariance	0.548[.579]	3.343[.035]	0.637[.530]

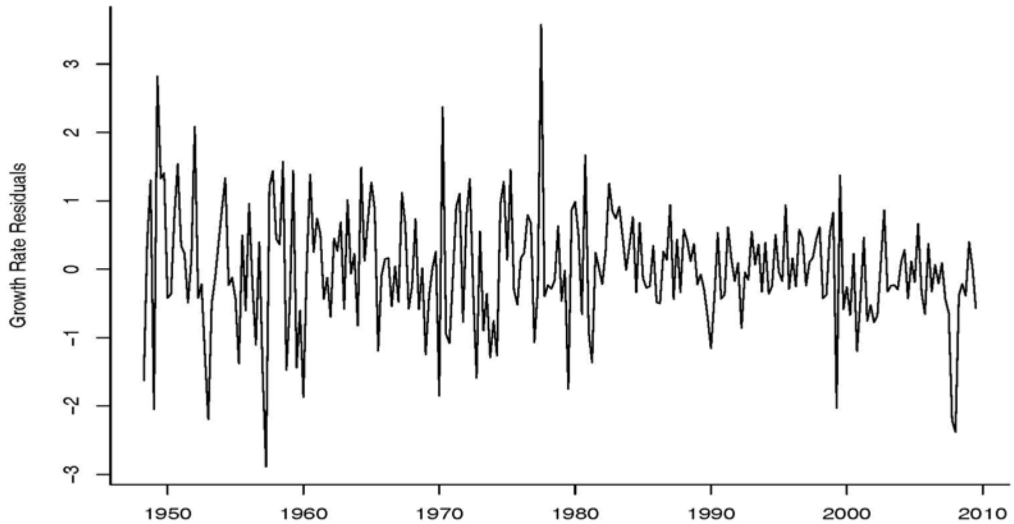


Fig. 7: N-VAR residuals ($\widehat{u}_t$)

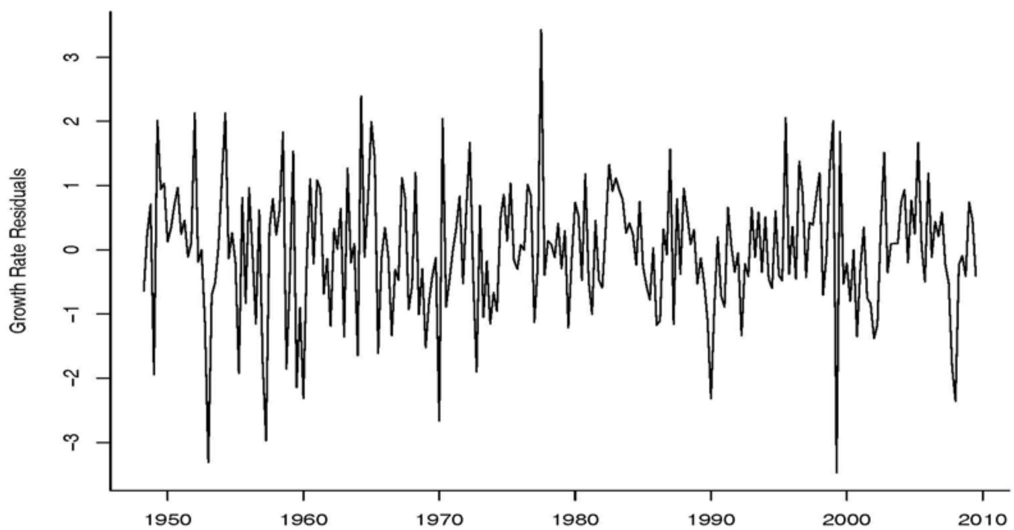


Fig. 8: St-VAR residuals ($\widehat{u}_t$)

This calls into question the widely accepted hypothesis known as the 'great moderation', claiming that the volatility of GDP growth during the period 1948-1983 is dramatically reduced for the 1984-2010, as indicated in figure 7. The above discussion suggests that the lower volatility arises as an inherent chance regularity stemming from $\{\mathbf{Z}_t,\ t{\in}\mathbb{N}\}$ when the latter is a Student's t Markov process. It represents a chance regularity naturally arising from the second order temporal dependence of the underlying process first noticed by Mandelbrot (1963):

> [5] "...large changes tend to be followed by large changes-of either sign – and small changes tend to be followed by small changes." (p. 418).

In summary, Student's t VAR(3) model constitutes a statistically adequate model which accounts for the chance regularities in data $\mathbf{Z}_0$. Hence, one can use this model to test the DSGE over-identifying restrictions:

$$H_0\text{: } \mathbf{G}(\boldsymbol{\theta},\boldsymbol{\varphi})=\mathbf{0}, \text{ vs. } H_1\text{: } \mathbf{G}(\boldsymbol{\theta},\boldsymbol{\varphi})\neq\mathbf{0}, \text{ for } \boldsymbol{\theta}\in\Theta,\ \boldsymbol{\varphi}\in\Phi,$$

knowing that the statistical adequacy of the model ensures that the actual error probabilities provide a close approximation to the nominal (assumed) ones. The relevant test is based on the likelihood ratio statistic:

$$\lambda_n(\mathbf{Z})=\frac{\max_{\boldsymbol{\varphi}\in\Phi} L(\boldsymbol{\varphi};\mathbf{Z})}{\max_{\boldsymbol{\theta}\in\Theta} L(\boldsymbol{\theta};\mathbf{Z})}=\frac{L(\widetilde{\boldsymbol{\varphi}};\mathbf{Z})}{L(\widehat{\boldsymbol{\theta}};\mathbf{Z})} \Rightarrow -2\ln\lambda_n(\mathbf{Z}) \overset{H_0}{\underset{a}{\backsim}} \chi^2(m). \tag{9}$$

For m=27, for $\alpha = .05$, $c_\alpha = 40.1$, the observed test statistic yields:

$$-2\ln\lambda_n(\mathbf{Z}_0)=5662.13[.000000000].$$

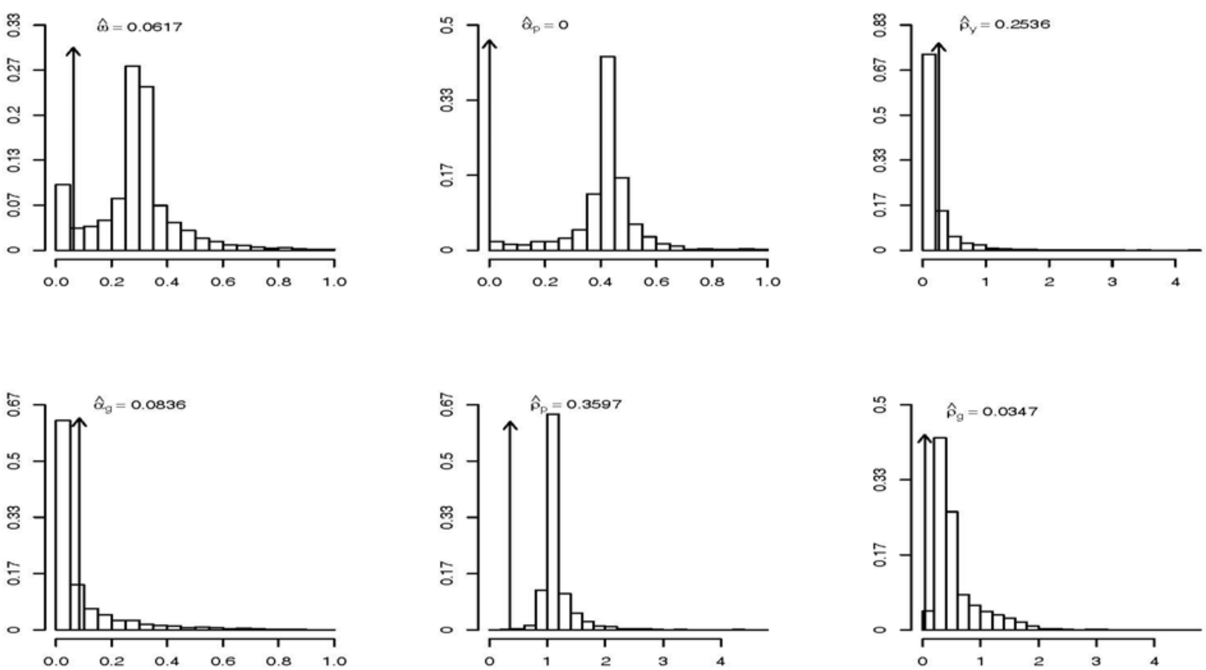
This result provides indisputably *strong evidence against* the DSGE model!

4.3 Sum-up assessment of DSGE modeling

4.3.1 Identification of the key structural parameters

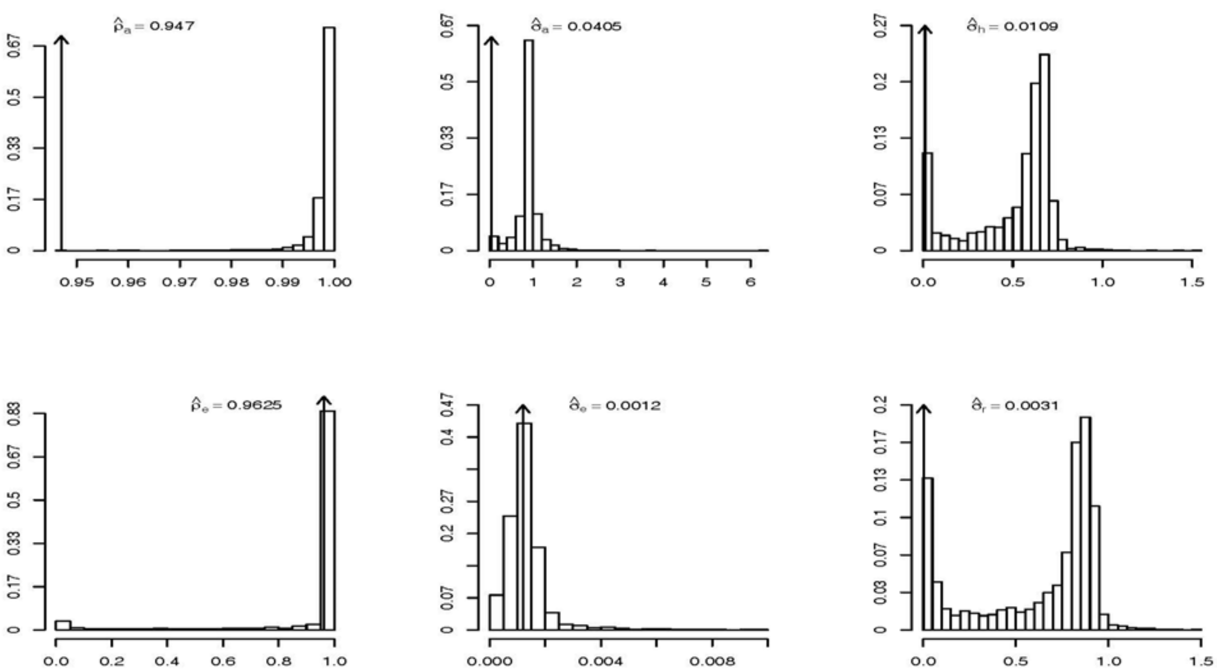
A crucial issue raised in the DSGE literature is the identification of the structural parameters; see Canova (2007). The problem is that often there is no direct way to relate the statistical ($\boldsymbol{\theta}$) to the structural parameters ($\boldsymbol{\varphi}$) because the implicit function $\mathbf{G}(\boldsymbol{\theta},\boldsymbol{\varphi}) = \mathbf{0}$ is not only highly non-linear, but it also involves algorithms like the Schur decomposition of the structural matrices involved.

An indirect way to probe the identification of the above DSGE model is to use the estimated statistical model, St-VAR(3;ν=3), whose statistical adequacy ensures that it accounts for the statistical regularities in the data, to generate faithful (true to the probabilistic structure of $\mathbf{Z}_0$) replications, say N, of the original data $\mathbf{Z}_0$.



The N simulated data series can then be used to estimate the structural parameters ($\boldsymbol{\varphi}$) using the original 'quantification' procedures. When the histogram of each $\widehat{\varphi}_i$, for i=1, 2, ..., p, is concentrated around a particular value, with a narrow interval of support, then φ_i can be considered *identifiable*. When the histogram associated with a particular $\widehat{\varphi}_i$, i=1, 2, ..., p, exhibits a large range of values, or/and multiple modes, indicate that the substantive parameter in question is not *identifiable*.

The 12 histograms below were generated using N=3000 replications of the original data of sample size n=252. Looking at these histograms two features stand out. First there are at least three parameters which are not identifiable. Second, out of the 9 identifiable parameters only 3 reported calibration values come close to the most likely value; for the other 6 parameters the calibrated value is very different. Increasing the number of replications N does not change the results. The t-invariance of statistical parameters is validated by the statistically adequate model, but the identification and constancy of the 'deep' DSGE parameters has been called into question by the above simulation exercise.



4.3.2 The gap between theory variables and data

It is well-known that any equations that result from of individual optimization would denote *intentions* (plans) in light of a range of hypothetical choices. The data measure what actually happened, the end result of numerous interactions among a multitude of agents over time. That is, what is observed in macro-data relates to the *adjustment* of the *realized* quantities and prices as they emerge from ever changing market conditions. This is exactly what is assumed away by DSGE modeling when equilibrium is imposed. As argued by Colander et al (2008):

"Any meaningful model of the macro economy must analyze not only the characteristics of the individuals but also the structure of their interactions." (p. 237)

4.3.3 Model validation vs. model calibration

Haavelmo (1940) prophetically warned against current DSGE strategies of producing models that can simulate 'realistic-looking data', arguing that this apparent 'degree of uniformity' can be illusory. This is because calibration is an unreliable procedure for ensuring that $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ accounts for the probabilistic structure of $\mathbf{Z}_0$. A more reliable way to appraise the validity of the statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ implicit in the DSGE model $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$ using simulation is to test the hypothesis:

Does $\mathbf{Z}_0 - \mathbf{Z}_0^s = \mathbf{U}$ constitute a realization of a white-noise process?

The theory-based Hodrick-Prescott (H-P) type filters and the equilibrium transformations are often ineffective in 'de-trending' and 'de-memorizing' the data, and as a result the transformed data often exhibit heterogeneity and dependence. This renders the usual estimators of the first two moments:

$$\widehat{Cov(\mathbf{Z})} = \frac{1}{n}\sum_{t=1}^{n}(\mathbf{Z}_t - \overline{\mathbf{Z}})(\mathbf{Z}_t - \overline{\mathbf{Z}})^\top, \ \overline{\mathbf{Z}} = \frac{1}{n}\sum_{t=1}^{n}\mathbf{Z}_t,$$

inconsistent, inducing a sizeable discrepancy between actual and nominal error probabilities for any inference based on such estimates; see Spanos and McGuirk (2001). In addition, moment-matching is not a reliable procedure to account for the regularities in the data; two random variables can have identical first four moments, but be completely different (Spanos, 1999).

4.3.4 Substantive vs. Statistical adequacy

Viewing the Lucas argument about abstraction and simplification from the error statistical perspective it is clear that it conflates substantive with statistical adequacy. There is nothing wrong with constructing a simple, abstract and idealized theory-model $\mathcal{M}_{\boldsymbol{\psi}}(\mathbf{z};\boldsymbol{\xi})$ aiming to capture key features of the phenomenon of interest, with a view to shed light on (understand, explain, forecast) economic phenomena of interest, as well as gain insight concerning alternative policies. The problem arises when the data $\mathbf{Z}_0$ are given a *subordinate* role, that of 'quantifying' $\mathcal{M}_{\boldsymbol{\psi}}(\mathbf{z};\boldsymbol{\xi})$ that (i) largely ignores the probabilistic structure of the data, (ii) employs unsound links between the model and the phenomenon of interest via $\mathbf{Z}_0$, and (iii) no testing of whether $\mathcal{M}_{\boldsymbol{\psi}}(\mathbf{z};\boldsymbol{\xi})$ does, indeed, capture the key features of the phenomenon of interest is carried out; see Spanos (2009).

Statistical misspecification is not an inevitable consequence of abstraction and simplification, but the result of ignoring the probabilistic structure of the data! When Kydland and Prescott (1991) argue:

"The reign of this system-of-equations macroeconomic approach was not long. One reason for its demise was the spectacular predictive failure of the approach." (p. 166)

it is clear that they have drawn the wrong lesson from the failure of the traditional macroeconometric models in the 1980s. Their predictive failure was primarily due to their substantive inadequacy. A weakness shared by today's DSGE models that also exhibit similar 'predictive failure'.

4.3.5 Poor forecasting performance

Typical examples of out-of-sample forecasting ability of both the DSGE and the Student's t VAR(3) models for 12 periods ahead [2003Q2-2006Q1; estimation period 1948Q1-2003Q1] is shown in figures 3-4 for GDP growth and inflation, with the actual data denoted by small circles.

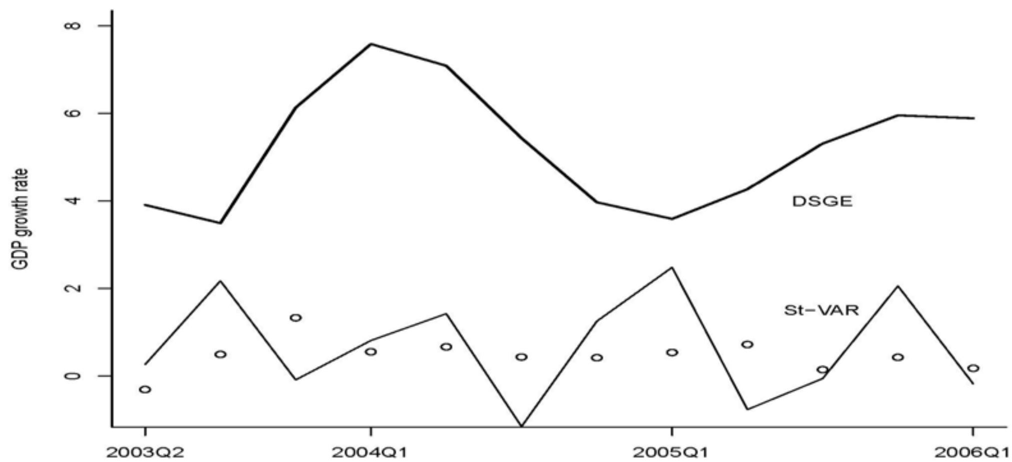


Fig. 3: Forecasting GDP growth

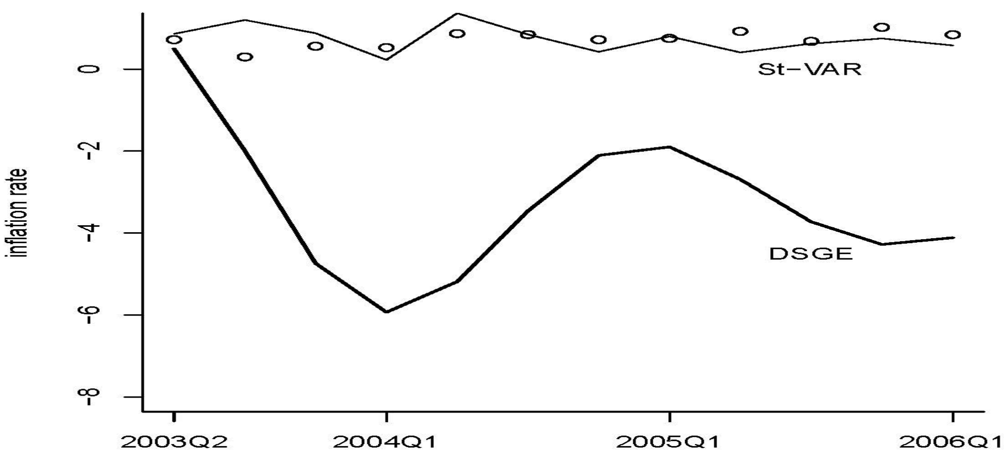


Fig. 4: Forecasting inflation

As can be seen, the performance of the DSGE is terrible in the sense that the prediction errors are both large and systematic [over/under prediction]; symptomatic of serious statistical inadequacy! The performance of the **St-VAR is excellent;** its prediction errors are both non-systematic and small! Interestingly, the poor forecasting performance of DSGE models is well-known, but it is rendered acceptable by invoking their *relative* performance:

"... we find that the benchmark estimated medium scale DSGE model forecasts inflation and GDP growth very poorly, *although statistical and judgemental forecasts do equally poorly*." (Edge and Gurkaynak, 2010, p. 209)

They failed to recognize that the poor forecasts were primarily due to the fact that they were based on statistically misspecified models; see tables 5 and 8.

4.3.6 Misleading impulse response analysis

The statistical inadequacy of the underlying statistical model also affects the reliability of its impulse response analysis, giving rise to misleading results about the reaction to exogenous shocks over time.

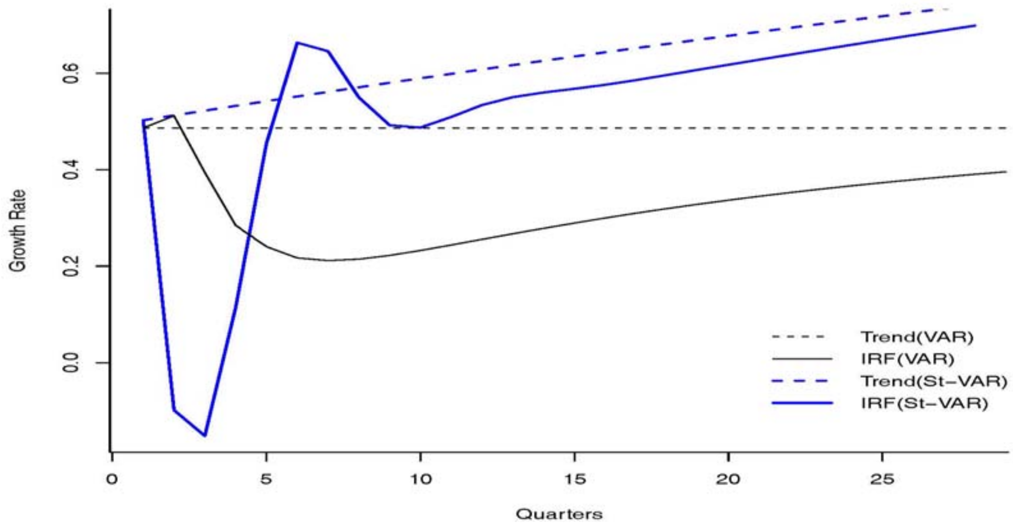


Fig. 5: 1% interest rate shock on GDP

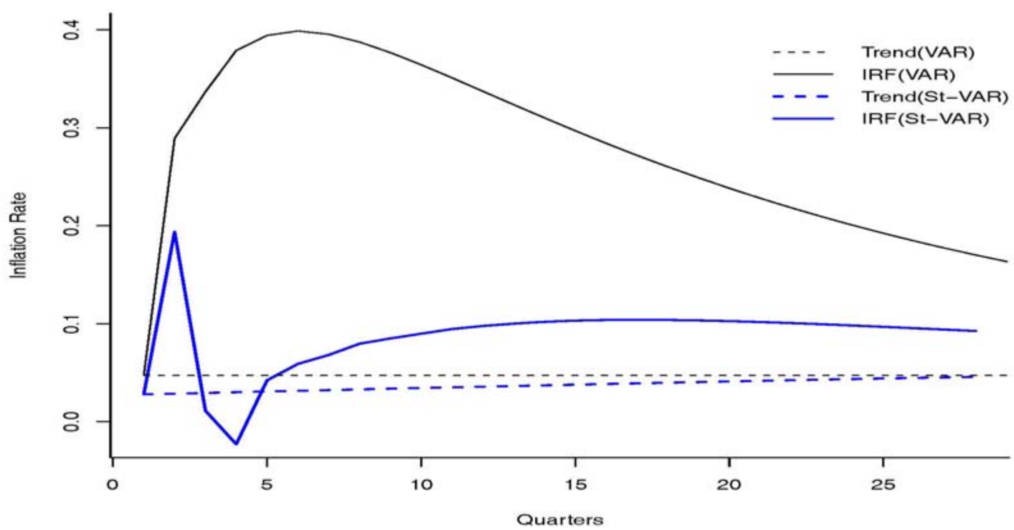


Fig. 6: 1% interest rate shock on inflation

Fig. 5 compares the impulse responses from a 1% increase in the interest rate (r_t) on per-capita real GDP from the Normal and Student's t VAR models. The heterogeneous St-VAR model produces a sharper decline and a sharper recovery in the growth rate of per-capita real GDP. This indicates stronger evidence for the effectiveness of the monetary policy. After some quarters of sharp decline, the growth rate for some time rises above the trend before falling below the trend again. But the effects produced by the stationary Normal VAR model is completely different. The growth rate smoothly falls and sluggishly recovers. The effects on the inflation rate are also significantly different in the two models, as shown in fig. 6.

5 Summary and conclusions

Empirical modeling and inference give rise to learning from data about phenomena of interest when applying reliable inference procedures using estimated statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ that are sufficiently adequate so that the actual error probabilities approximate closely the assumed (nominal) ones. Imposing the theory on the data often leads to an impasse, since the estimated model is often both statistically and substantively inadequate, rendering any proposed substantive respecifications of the original structural model questionable; the respecified model is declared 'better' on the basis of untrustworthy evidence! This is because no evidence for or against a substantive claim can be secured on the basis of a statistically misspecified model.

The error-statistical way to address this impasse is to separate, *ab initio*, the substantive $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$ from the statistical premises $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ and establish statistical adequacy before posing any substantive questions of interest. This is achieved by viewing $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ as a parameterization of the process $\{\mathbf{Z}_t,\ t\in\mathbb{N}\}$ underlying data $\mathbf{Z}_0$, that (parametrically) *nests* $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$ via $\mathbf{G}(\boldsymbol{\theta},\boldsymbol{\varphi})=\mathbf{0}$.

DSGE modeling exemplifies the Ricardian vice in theory-driven modeling. 'Empirically plausible calibration or estimation which fits the main features of the macroeconomic time series' is much too unsound a link to reality. Theory-driven vs. data-driven, realistic vs. unrealistic and policy-oriented vs. non-policy oriented models, are false dilemmas! An estimated DSGE model $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$ whose statistical premises $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ are misspecified constitutes a poor and totally unreliable basis for any form of inference, including appraising substantive adequacy, forecasting, policy simulations and investigating the dynamic transmission mechanisms of the real economy. Amazingly, Haavelmo (1940, 1943, 1944) largely anticipated most of these potential weaknesses [a]-[e] mentioned above; see Spanos (2013).

Confronting a DSGE model $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$ with reliable evidence in the form of a statistically adequate $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ [Student's t VAR(3)] strongly rejects $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$. In light of the above discussion, a way forward for DSGE modeling is to:

(a) bridge the gap between theory and data a lot more carefully,

(b) avoid imposing theory-based restrictions on the data at the outset, including the H-P filtering of the data,

(c) account for the probabilistic structure of the data by securing the statistical adequacy $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$, before any inferences are drawn.

In the meantime, the estimated Student's t VAR(3) can play a crucial role in:

(i) guiding the search for better theories by demarcating 'what there is to explain',

(ii) generating more reliable short-term forecasts and policy simulations until an empirically valid DSGE model is built.

References

[1] Blaug, M. (1958), *Ricardian Economics: a Historical Study*, Yale University Press, CT.

[2] Burns, A. F. and W. C. Mitchell (1946), *Measuring the Business Cycles*, National Bureau of Economic Research, NY.

[3] Cairnes, J. E. (1888), *The Character and Logical Method of Political Economy*, Reprints of Economic Classics, 1965, Augustus, M. Kelley, NY.

[4] Canova, F. (2007), *Methods for Applied Maeconomic Research*, Princeton Unversity Press, NJ.

[5] Colander, D. P. Howitt, A. Kirman, A. Leijonhufvud and P. Mehrling (2008), "Beyond DSGE Models: Toward an Empirically Based Macroeconomics," *The American Economic Review*, **98**: 236-240.

[6] Cooper, R. L. (1972), "The predictive performance of quarterly econometric models of the United States," in *Econometric Models of Cyclical Behavior*, ed. by Hickman, B. G., Columbia University Press, NY.

[7] DeJong, D. N. and C. Dave (2011), *Structural Macroeconometrics*, 2nd ed., Princeton Unversity Press, NJ.

[8] Edge, R. M. and R. S. Gurkaynak (2010), "How Useful are Estimated DSGE Model Forecasts for Central Bankers?", *Brookings Papers*, 2011, pp. 209-244.

[9] Favero, C. A. (2001), *Applied Macroeconometrics*, Oxford University Press, Oxford.

[10] Frisch, R. (1933), "Propagation problems and Impulse Problems in Dynamic Economics," in Economic Essays in Honour of Gustav Cassel, Allen and Unwin, London.

[11] Frisch, R. (1934), *Statistical Confluence Analysis by Means of Complete Regression Schemes*, Universitetets Okonomiske Institutt, Oslo.

[12] Fromm, G. and L. R. Klein (ed.) (1975), *Brookings Model: Perspective and Recent Developments*, Elsevier, North-Holland.

[13] Granger, C. W. J. (1990), "Where are the Controversies in Econometric Methodology," pp. 1-23 in Granger, C. W. J. (ed.) *Modelling Economic Series*, Clarendon Press, Oxford.

[14] Granger, C. W. J. and P. Newbold (1986), *Forecasting Economic Time Series*, 2nd, ed., Academic Press, London.

[15] Greene, W. H. (2011), *Econometric Analysis*, 7th edition, Pearson/Prentice Hall, NJ.

[16] Haavelmo, T. (1940), "The Inadequacy of Testing Dynamic Theory by Comparing Theoretical Solutions and Observed Cycles," *Econometrica*, **8**: 312-321.

[17] Haavelmo, T. (1943), "The Statistical Implications of a System of Simultaneous Equations," *Econometrica,* **11**: 1-12.

[18] Haavelmo, T. (1944), "The Probability Approach in Econometrics", *Econometrica*, **12**, supplement, 1-115.

[19] Haavelmo, T. (1989), "Econometrics and the Welfare State," Nobel lecture, reprinted in *The American Economic Review*, **87**: 13-15.

[20] Haberler, G. (1937), *Prosperity and Depression*, League of Nations, Geneva.

[21] Heer, B. and A. Maussner (2009), *Dynamic General Equilibrium Modeling*, 2nd ed., Springer, NY.

[22] Hempel, C. (1965), *Aspects of Scientific Explanation*, Free Press, NY.

[23] Hendry, D. F. and M. S. Morgan (eds.) (1995), *The Foundations of Econometric Analysis*, Cambridge University Press, Cambridge.

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

[25] Keynes, J. M. (1939), "Professor Tinbergen's Method," *Economic Journal*, **49**: 558-568.

[26] Kirman, A. (1992), Whom or what does the representative individual represent?, *Journal of Economic Perspectives*, **6**: 117–36.

[27] Klein, P. (2000), "Using the generalized schur form to solve a multivariate linear rational expectations model," *Journal of Economic Dynamics and Control*, 24:1405–1423.

[28] Klein, D. R. and A. S. Goldberger (1955), *Econometric Model of the United States, 1929-52*, Elsevier, North-Holland.

[29] Kydland, F. E. and E. C. Prescott (1991), "The Econometrics of the General Equilibrium Approach to the Business Cycles," *The Scandinavian Journal of Economics*, 93: 161-178.

[30] Lucas, R. E. (1976), "Econometric Policy Evaluation: a Critique," pp. 19-46 of *The Phillips Curve and Labour Markets*, ed. by K. Brunner and A. M. Metzer, Carnegie-Rochester Conference on Public Policy I, North-Holland, Amsterdam.

[31] Lucas, R. E. (1980), "Methods and Problems in Business Cycle Theory," *Journal of Money, Credit and Banking*, 12: 696-715.

[32] Lucas, R. E. (1987), *Models of Business Cycles*, Blackwell, Oxford.

[33] Lucas, R. E. and T. J. Sargent (1981), Rational Expectations and Econometric Practice, George Allen & Unwin, London.

[34] Mandelbrot, B. (1963) "The variation of certain speculative prices," *Journal of Business*, pp. 394-419.

[35] Mayo, D. G. and A. Spanos (2004), "Methodology in Practice: Statistical Misspecification Testing", Philosophy of Science, 71: 1007-1025.

[36] Mayo, D. G. and Spanos, A. (2006), "Severe testing as a basic concept in a Neyman–Pearson philosophy of induction," *British Journal for the Philosophy of Science*, 57: 323–57.

[37] McCarthy, M. D. (1972), *The Wharton quarterly econometric forecasting model Mark III*, Economic Research Unit, Department of Economics, Wharton School of Finance and Commerce, University of Pennsylvania.

[38] Mitchell, W. C. (1913), *Business Cycles and their Causes*, California University Memoirs vol. 3, Berkeley.

[39] Prescott, E. C. (1986), "Theory Ahead of Business Cycle Measurement," Federal Reserve Bank of Minneapolis, Quarterly Review, 10: 9-22.

[40] Rao, C. R. and S. K. Mitra (1972), *Generalized Inverse of Matrices and its Applications*, Wiley, NY.

[41] Robbins, L. (1935), *An Essay on the Nature and Significance of Economic Science*, 2nd ed., McMillan, London.

[42] Spanos, A., (1986), *Statistical Foundations of Econometric Modelling*, Cambridge University Press, Cambridge.

[43] Spanos, A. (1990), "The Simultaneous Equations Model revisited: statistical adequacy and identification", *Journal of Econometrics*, **44**, 87-108.

[44] Spanos, A. (1994), "On Modeling Heteroskedasticity: the Student's t and Elliptical Linear Regression Models," *Econometric Theory*, **10**: 286-315.

[45] Spanos, A. (1995), "On theory testing in Econometrics: modeling with nonexperimental data", *Journal of Econometrics,* **67**: 189-226.

[46] Spanos, A. (1999), *Probability Theory and Statistical Inference: econometric modeling with observational data*, Cambridge University Press, Cambridge.

[47] Spanos, A. (2006a), "Econometrics in Retrospect and Prospect," pp. 3-58 in Mills, T.C. and K. Patterson, *New Palgrave Handbook of Econometrics*, vol. 1, MacMillan, London.

[48] Spanos, A. (2006b), "Where Do Statistical Models Come From? Revisiting the Problem of Specification," pp. 98-119 in *Optimality: The Second Erich L. Lehmann Symposium*, edited by J. Rojo, Lecture Notes-Monograph Series, vol. 49, Institute of Mathematical Statistics.

[49] Spanos, A. (2006c), "Revisiting the omitted variables argument: substantive vs. statistical adequacy," *Journal of Economic Methodology*, **13**: 179–218.

[50] Spanos, A. (2007), "Curve-Fitting, the Reliability of Inductive Inference and the Error-Statistical Approach," *Philosophy of Science*, **74**: 1046–1066.

[51] Spanos, A. (2009), "The Pre-Eminence of Theory versus the European CVAR Perspective in Macroeconometric Modeling," *Economics: The Open-Access, Open-Assessment E-Journal*, Vol. 3, 2009-10. http://www.economics-ejournal.org/economics/journalarticles/2009-10.

[52] Spanos, A. (2010a), "Akaike-type Criteria and the Reliability of Inference: Model Selection vs. Statistical Model Specification," *Journal of Econometrics,* **158**: 204-220.

[53] Spanos, A. (2010b), "Theory Testing in Economics and the Error Statistical Perspective," pp. 202-246 in *Error and Inference*, edited by Mayo, D. G. and A. Spanos (2010).

[54] Spanos, A. (2013), "Revisiting Haavelmo's Structural Econometrics: Bridging the Gap between Theory and Data,", Virginia Tech, working paper.

[55] Spanos, A. and A. McGuirk (2001), "The Model Specification Problem from a Probabilistic Reduction Perspective," *Journal of the American Agricultural Association*, **83**: 1168-1176.

[56] Tinbergen, J. (1939), *Statistical Testing of the Business-Cycle Theories*, Vol. I-II, League of Nations, Geneva.

Appendix- Miscellaneous results

5.1 Restricted state-space formulation matrices

$$\mathbf{A}=\begin{bmatrix} 0 & -1 & 0 & 0 & 0 & 1 & 1-\alpha_x \\ 0 & 0 & 0 & 0 & \psi & \beta(1-\alpha_x) & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -\rho_\pi & -\rho_g & -\rho_z & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \mathbf{B}=\begin{bmatrix} 0 & 0 & 0 & 0 & -\alpha_x & 0 & 1 \\ 0 & 0 & -\beta\alpha_x & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & \rho_r & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{C}=\begin{bmatrix} (1-\omega)(1-\rho_a) & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \omega & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \mathbf{P}=\begin{bmatrix} \rho_a & 0 & 0 & 0 \\ 0 & \rho_e & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

5.2 Derivation of Reduced Form Structural model

The system of equations (5) can be decomposed into two subsystems by defining a vector of the observable variables $\mathbf{d_t}=\begin{bmatrix}\widehat{y}_t & \widehat{p}_t & \widehat{r}_t\end{bmatrix}^\top$ and a vector of the unobservable variables $\mathbf{Y_t}=\begin{bmatrix}\widehat{y}_{ht} & \widehat{g}_t\end{bmatrix}^\top$ as follows:

$$\mathbf{d}_t = \mathbf{D}\mathbf{Y}_{t-1} + \mathbf{E}\mathbf{d}_{t-1} + \mathbf{F}\mathbf{v}_t \tag{10}$$

$$\mathbf{Y}_t = \mathbf{G}\mathbf{Y}_{t-1} + \mathbf{H}\mathbf{d}_{t-1} + \mathbf{K}\mathbf{v}_t \tag{11}$$

where $\mathbf{v}_t = \mathbf{P}\mathbf{v}_{t-1} + \boldsymbol{\varepsilon}_t$ and $\mathbf{D}, \mathbf{E}, \mathbf{F}, \mathbf{G}, \mathbf{H}, \mathbf{K}$ are formed by the partitioning of $\mathbf{A}_1(\psi)$ as follows:

$$\mathbf{A}_1(\psi)_{(9\times 9)} = \begin{bmatrix} \mathbf{E}_{(3\times 3)} & \mathbf{D} & \mathbf{F} \\ \mathbf{H} & \mathbf{G}_{(2\times 2)} & \mathbf{K} \\ \mathbf{0} & \mathbf{0} & \mathbf{P}_{4\times 4} \end{bmatrix}$$

Since $\mathbf{D}$ is a rectangular matrix of dimension (3×2), usual $\mathbf{D}^{-1}$ does not exist. So the generalized inverse is used to eliminate $\mathbf{Y_t}$ from the system (10) and (11), which yields (whenever the usual inverse of the matrix does not exist, the generalized inverse is used following Rao and Mitra (1971). The generalized inverse is the same as the regular inverse when the inverse of the matrix exists):

$$\mathbf{d}_t = \left[\mathbf{D}\mathbf{G}\mathbf{D}^{-1} + \mathbf{E}\right]\mathbf{d}_{t-1} + \mathbf{D}(\mathbf{H} - \mathbf{G}\mathbf{D}^{-1}\mathbf{E})\mathbf{d}_{t-2} + \mathbf{e}_t \tag{12}$$

$$\mathbf{e}_t = \mathbf{F}\mathbf{v}_t + \mathbf{D}(\mathbf{K} - \mathbf{G}\mathbf{D}^{-1}\mathbf{F})\mathbf{v}_{t-1} \tag{13}$$

$$\mathbf{v_t} = \mathbf{P}\mathbf{v_{t-1}} + \boldsymbol{\varepsilon}_t \tag{14}$$

Using (12), (13) and (14), $\mathbf{v}_t$ can be eliminated to yield:

$$\mathbf{d}_t = \boldsymbol{\Psi}_1\mathbf{d}_{t-1} + \boldsymbol{\Psi}_2\mathbf{d}_{t-2} + \mathbf{u}_t \tag{15}$$

$$\mathbf{u}_t = \boldsymbol{\Psi}_3\boldsymbol{\varepsilon}_t \quad \boldsymbol{\Psi}_1 = \mathbf{D}\mathbf{G}\mathbf{D}^{-1} + \mathbf{E}, \boldsymbol{\Psi}_2 = \mathbf{D}(\mathbf{H} - \mathbf{G}\mathbf{D}^{-1}\mathbf{E}),\ \boldsymbol{\Psi}_{3(3\times 4)} = \left[\boldsymbol{\Lambda} - \mathbf{I}\right]^{-1}\boldsymbol{\Lambda}\mathbf{F},$$
$$\boldsymbol{\Lambda}_{3\times 3} = \mathbf{F}\mathbf{P}(\mathbf{F}\mathbf{P} + \mathbf{D}(\mathbf{K} - \mathbf{G}\mathbf{D}^{-1}\mathbf{F}))^{-1} + \mathbf{D}(\mathbf{K} - \mathbf{G}\mathbf{D}^{-1}\mathbf{F})(\mathbf{F}\mathbf{P} + \mathbf{D}(\mathbf{K} - \mathbf{G}\mathbf{D}^{-1}\mathbf{F}))^{-1}$$
$$(\mathbf{u}_t|\mathbf{d}_{t-1}, \mathbf{d}_{t-2}) \sim \mathsf{N}(0, \boldsymbol{\Psi_3}\mathbf{V}\boldsymbol{\Psi_3}^{\top}).$$

Since, $\mathbf{d}_t = \mathbf{Z}_t - \mathbf{z}$, where $\mathbf{Z}_t = [\ln(y_t), \ln(p_t), \ln(r_t)]^{\top}$ and $\mathbf{z} = [\ln(\overline{y}), \ln(\overline{p}), \ln(\overline{r})]^{\top}$ is steady state, equation (15) can be written as:

$$\mathbf{Z}_t = \boldsymbol{\psi}_0 + \boldsymbol{\Psi}_1\mathbf{Z}_{t-1} + \boldsymbol{\Psi}_2\mathbf{Z}_{t-2} + \mathbf{u}_t \tag{16}$$

where, the condition $\boldsymbol{\psi}_0 = (\mathbf{I} - \boldsymbol{\Psi_1} - \boldsymbol{\Psi_2})\mathbf{z}$ can be used to identify $\mathbf{z}$ once 7 is estimated. The constant steady state assumption in the DSGE model is disproved by data in this paper. By relaxing all the structural restrictions imposed in (23), the statistical model in the form of Normal VAR(2) in Table 6 is obtained.

5.3 Multivariate Student's t

For $\mathbf{X} \backsim \mathsf{St}(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$,where $\mathbf{X} : p \times 1$, the joint density function is:

$$f(\mathbf{x}; \boldsymbol{\varphi}) \ = (\nu\pi)^{-\frac{p}{2}}\frac{\Gamma(\frac{p+\nu}{2})}{\Gamma(\frac{\nu}{2})}(\det\boldsymbol{\Sigma})^{-\frac{1}{2}}\{1 + \tfrac{1}{\nu}(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\}^{-(\frac{p+\nu}{2})},$$

where $\boldsymbol{\varphi} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ $E(\mathbf{X}) = \boldsymbol{\mu}$, $Var(\mathbf{X}) = \frac{\nu}{\nu-2}\boldsymbol{\Sigma}$.

Student's t VAR (St-VAR) Model

Let $\{\mathbf{Z}_t, t = 1, 2, \ldots\}$ be a vector Student's t with ν df, Markov(l) and stationary process. The joint distribution of $\mathbf{X}_t := (\mathbf{Z}_t, \mathbf{Z}_{t-1}, \ldots, \mathbf{Z}_{t-l})$ is denoted by:

$$\mathbf{X}_t \sim \mathsf{St}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu\right)$$

$$\mathbf{X}_t = \begin{bmatrix} \mathbf{Z}_t \\ \mathbf{Z}_{t-1} \\ \mathbf{Z}_{t-2} \\ \dots \\ \mathbf{Z}_{t-l} \end{bmatrix} \sim \mathsf{St}\left(\begin{bmatrix} \boldsymbol{\mu}_z \\ \boldsymbol{\mu}_z \\ \boldsymbol{\mu}_z \\ .. \\ \boldsymbol{\mu}_z \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{13} & \dots & \boldsymbol{\Sigma}_{1l+1} \\ \boldsymbol{\Sigma}_{12}^\top & \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \dots & \boldsymbol{\Sigma}_{1l} \\ \boldsymbol{\Sigma}_{13}^\top & \boldsymbol{\Sigma}_{12}^\top & \boldsymbol{\Sigma}_{11} & \dots & \boldsymbol{\Sigma}_{1l-1} \\ \dots & \dots & \dots & \dots & \dots \\ \boldsymbol{\Sigma}_{1l+1}^\top & \boldsymbol{\Sigma}_{1l}^\top & \boldsymbol{\Sigma}_{1l-1}^\top & \dots & \boldsymbol{\Sigma}_{11} \end{bmatrix}; \nu \right)$$

where $\mathbf{Z}_t$: $(k \times 1)$, $\boldsymbol{\Sigma}_{ij}$: $(k \times k)$, $\boldsymbol{\mu}_z$: $(k \times 1)$, $\boldsymbol{\mu}$: $(p \times 1)$, $\boldsymbol{\Sigma}$: $(p \times p)$ $p{=}(l{+}1)k$-number of variables in $\mathbf{X}_t$, k-number of variables in $\mathbf{Z}_t$, l-number of lags.

Joint, Conditional and Marginal Distributions

Let the vectors $\mathbf{X}_t$ and $\boldsymbol{\mu}$, and the matrix $\boldsymbol{\Sigma}$ are partitioned as follows:

$$\mathbf{X}_t{=}\begin{bmatrix} \mathbf{Z}_t(k \times 1) \\ \mathbf{Z}^0_{t-1}(lk \times 1) \end{bmatrix}, \quad \boldsymbol{\mu}{=}\begin{bmatrix} \boldsymbol{\mu}_z(k \times 1) \\ \boldsymbol{\mu}_{lk}(lk \times 1) \end{bmatrix}, \quad \boldsymbol{\Sigma}{=}\begin{bmatrix} \boldsymbol{\Sigma}_{11}(k \times k) & \boldsymbol{\Sigma}_{12}(k \times lk) \\ \boldsymbol{\Sigma}_{12}^\top(lk \times k) & \mathbf{Q}(lk \times lk) \end{bmatrix}$$

Here, $\boldsymbol{\mu}_{lk}(lk \times 1)$ is a vector of lk $\boldsymbol{\mu}_z$'s. Now, the joint, the conditional and the marginal distributions for all $t{\in}\mathbb{N}$ are denoted by:

$$D(\mathbf{Z}_t, \mathbf{Z}^0_{t-1}; \boldsymbol{\theta}){=}D(\mathbf{Z}_t|\mathbf{Z}^0_{t-1}; \boldsymbol{\theta}_1)D(\mathbf{Z}^0_{t-1}; \boldsymbol{\theta}_2) \sim \mathsf{St}(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$$

$$D(\mathbf{Z}_t|\mathbf{Z}^0_{t-1}; \boldsymbol{\theta}_1){\sim}\mathsf{St}(\mathbf{a}_0 + \mathbf{A}^\top\mathbf{Z}^0_{t-1}, \boldsymbol{\Omega} q(\mathbf{Z}^0_{t-1}); \nu + lk) \quad D(\mathbf{Z}^0_{t-1}; \boldsymbol{\theta}_2){\sim}\mathsf{St}(\boldsymbol{\mu}_{lk}, \mathbf{Q}; \nu)$$

$$q(\mathbf{Z}^0_{t-1}){=} \left[1{+}\tfrac{1}{\nu}(\mathbf{Z}^0_{t-1}{-}\boldsymbol{\mu}_{lk})^\top \mathbf{Q}^{-1}(\mathbf{Z}^0_{t-1}{-}\boldsymbol{\mu}_{lk})\right] \quad \mathbf{A}^\top{=}\boldsymbol{\Sigma}_{12}\mathbf{Q}^{-1}$$

$$\mathbf{a}_0{=}\boldsymbol{\mu}_z{-}\mathbf{A}^\top\boldsymbol{\mu}_{lk}, \quad \boldsymbol{\Omega}{=}\boldsymbol{\Sigma}_{11}{-}\boldsymbol{\Sigma}_{12}\mathbf{Q}^{-1}\boldsymbol{\Sigma}_{12}^\top$$

$\mathbf{Z}^0_{t-1}{:=}(\mathbf{Z}_{t-1}, \dots, \mathbf{Z}_{t-l})$, $\boldsymbol{\theta}_1{=}\{\mathbf{a}_0, \mathbf{A}, \boldsymbol{\Omega}, \mathbf{Q}, \boldsymbol{\mu}\}$, $\boldsymbol{\theta}_2{=}\{\boldsymbol{\mu}, \mathbf{Q}\}$. The lack of variation freeness (Spanos, 1994) calls for defining the likelihood function in terms of the joint distribution, but reparameterized in terms of the conditional and marginal distribution parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, respectively.

This can be easily extended to a heterogeneous St-VAR model where the mean is assume to be: $\boldsymbol{\mu}_z(t) = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t + \boldsymbol{\mu}_2 t^2$. This makes the autoregressive function a quadratic function of t: $\mathbf{a}_0{=}\boldsymbol{\mu}_z(t){-}\mathbf{A}_1^T\boldsymbol{\mu}_z(t{-}1){-}\mathbf{A}_2^T\boldsymbol{\mu}_z(t{-}2){-}\mathbf{A}_3^T\boldsymbol{\mu}_z(t{-}3){=}\boldsymbol{\delta}_0 + \boldsymbol{\delta}_1 t + \boldsymbol{\delta}_2 t^2$. One important aspect of this model is that although heterogeneity is imposed only in mean of the joint distribution, both mean and variance-covariance of the conditional distribution are heterogeneous (i.e. functions of t).

5.4 Software

R software is produced to estimate the St-VAR model using maximum likelihood method for a given number of variables and lag length. The function in R is:

```
StVAR(Z,p,df,maxit,meth,hessian, init, trend)
```

`Z`: data matrix with observations in rows, `p`: number of lags, `df`: degrees of freedom, `maxit`: Number of iteration to be done for optimization,

`meth`: Any optimiation method used in optim function in R, `hessian`: TRUE/FALSE, `init`: initial values, `trend`: Q, L or C for quadratic, linear or constant.

The function `StVAR(.)` **returns** following inference results:

The estimated coefficients of autoregressive and autoskedastic functions with standard errors and p-values, Conditional variance covariance, Fitted values, Residuals, Likelihood value, M-S testing results.