A Tale of Fat Tails*

Chetan Dave† NYUAD
Samreen Malik‡ NYUAD

September 27, 2016

Abstract

We document the extent to which major macroeconomic series, used to inform linear DSGE models, can be characterized by power laws whose indices we estimate via maximum likelihood. Assuming data follow a linear recursion with multiplicative noise, low estimated indices suggest fat tails. We then ask whether standard DSGE models under constant gain learning can replicate those fat tails by an appropriate increase in the estimated gain and without much change in the transmission mechanism of shocks. We find that is largely the case via implementation of a minimum distance estimation method that eschews any allegiance to distributional assumptions or the matching of particular moments.

Keywords: Adaptive learning, DSGE models, Fat tails, Power law

*We thank Jess Benhabib, Christian Haefke, Jean Imbs, John Leahy, Max Mihm, Romain Ranciere, Gilles Saint-Paul and seminar participants at the EEA-ESEM congress, University of Winnipeg, Indian Statistical Institute and NYU Shanghai for helpful comments and suggestions. The usual disclaimer applies.

†cdave@nyu.edu
‡samreen.malik@nyu.edu

1. Introduction

The ability of linear dynamic stochastic general equilibrium (DSGE) models to adequately account for macroeconomic fluctuations has come under scrutiny in light of the Great Recession. Such large but rare events manifest themselves in the form of fat tails for data that are usually employed in standard Gaussian empirical DSGE modeling under rational expectations (RE). However, in the absence of additional assumptions on the stochastic nature of innovations, standard DSGE models are unable to replicate observed large fluctuations. We show that without altering the nature of structural shocks, a DSGE model under adaptive learning endogenously delivers model dynamics that better replicate observed fat tails.

Existing DSGE analyses have explored at least three avenues to model large macroeconomic fluctuations. The first avenue replaces the assumption of Normally distributed innovations with a fat-tailed specification (e.g. a Students'-t or Laplace). Fat tails in the distribution of innovations allow for a higher probability that a large shock occurs and works its way through standard transmission mechanisms. The second avenue replaces the assumption of a constant variance for structural innovations with exogenous stochastic volatility specifications. The idea is that if one introduces exogenous volatility into a DSGE shock specification then macroeconomic variables will also exhibit the sort of volatility associated with rare but large deviations. The third avenue introduces time variation in the structural parameters of a model which in turn generates time or state dependent responses of economies to an otherwise constant variance shock process. All three avenues modify models so that exogenous sources of volatility are introduced in order to match observed volatility. Our analysis is closest to the third avenue and presents an endogenous channel via stochastic gradient constant gain (SGCG) adaptive learning that delivers fat tails for endogenous variables in an otherwise standard model.

Adaptive learning is increasingly used by macroeconomists to bridge data-model gaps. Seminal work on statistical learning (Sargent, 1993 and Evans and Honkapohja, 2001), additional insights from a similar literature (Gaspar et al., 2006, Orphanides and Williams, 2004, Milani, 2007, Deak et. al., 2015, Massaro, 2013 and De Grauwe, 2012), and experimental evidence on the importance of the learning process in accounting for business cycle fluctuations and volatility (Duffy, 2012, Bao, Hommes, Sonnemans and Tuinstra, 2012, Jaimovich and Rebelo, 2007, Adam and Woodford, 2012) provide strong support that adaptive learning is a reasonable alternative to the standard RE DSGE setting. In contrast to the standard RE DSGE model, in which agents know the true stochastic process of an economy, under adaptive learning agents revise their forecast rules in response to incoming data so as to ascertain that stochastic process over time.[1] This difference in how expectations are formed influences model dynamics. In particular, under RE, model dynamics are described by a fixed coefficient vector autoregression (VAR). Under SGCG learning however, model dynamics are described by a linear recursion with multiplicative and additive noise (LRMN) instead of a fixed or time varying VAR. A LRMN, written as

$$X_t = \Phi_t X_{t-1} + \varepsilon_t, \tag{1}$$

[1] In the limit as the constant gain (g) of a SGCG learning process tends to zero, a model economy approaches the rational expectations solution (see Evans and Honkapohja, 2001).

has a stationary distribution for X_t different than that of a VAR due to the interplay between the *stochastic* multiplicative term (Φ_t) and the stochastic additive term (ε_t).[2] As a function of the nature of the interplay, the applied mathematics literature shows that the tail of the stationary distribution of X_t can be fatter than that of a Normal distribution (e.g. Kesten, 1973). This implies that X_t can take on extreme values with a higher probability than under a Normal distribution and thus this equation forms an alternative lens with which to view data and an associated model economy under SGCG learning. In a sense the intuition of a LRMN system is as follows: as grains of sand pile up into a dune, at some point a grain falls that shifts the dune dramatically, and this dramatic movement occurs with some regularity.[3]

Since under SGCG learning the same DSGE model is written as a LRMN, one needs to change the underlying assumption on the data generating process (DGP) from a fixed coefficient VAR to a LRMN. Under this new assumption, the tail of the stationary distribution of data (Y_t) can be fat. We measure the thickness of the tail by estimating the tail index p under the assumption that $Y_t \sim Y^{-p}$ (a power law), since under SGCG learning model variables are distributed similarly. We also provide a more conventional measure of kurtosis to establish that data that enter a DSGE empirical exercise are not necessarily Normal.

We find that data exhibit characteristics consistent with a LRMN assumption on the DGP. In particular, the usual data employed in DSGE models have fatter tails than would be warranted under a Normality assumption. We implement a minimum distance estimation exercise that allows us to jointly estimate model parameters including the constant gain (g). Were our estimates of g small or near zero then model dynamics would approximate those of a RE DSGE model. However, we find that estimates of g are non-zero and in fact higher than in the current literature. A finding of a higher constant gain does not violate any theoretical or empirical requirement that g be near zero (so a learning model is in a small vicinity of its RE solution). Moreover, our estimates of this gain are not inconsistent with those found by Malmendier and Nagel (2016) using structural micro-econometric analyses of survey data. We next conduct a simulation exercise in which we ask how the tail index of model variables varies around baseline parameter values as we vary only g; we find that as g increases the tail index of model variables declines. Further we show that a model under RE with Normally distributed innovations, or a model with fat-tailed distributions for innovations, is not able to come as close as a DSGE model under SGCG learning in terms of being able to replicate the fat tails observed in data.

These empirical and simulation results allow us to establish a central intuition, and therefore our key contribution: given that a larger g reflects a shorter memory (learning horizon), as g rises, macroeconomic variables are more likely to visit extreme values (deep recessions and booms), simply because agents do not remember as much of history as they could and therefore are bound to repeat it. Overall, we are able to show that without departing fundamentally from the standard DSGE model, SGCG learning is enough to account for observed fat tails.

[2] Our definitions for the process $X_t = \Phi_t X_{t-1} + \varepsilon_t$ are as follows. If Φ_t is a constant matrix Φ then we call it a fixed coefficient VAR. If Φ_t varies over time in a deterministic manner we call it a variable coefficient VAR. If Φ_t is itself a stochastic process then we refer to the equation as a LRMN.

[3] For similar intuition relating to the notion of self-organized criticality that a LRMN represents, see Blume et al. (2010).

Having established the model and described related literature in Sections 2 and 3, we set out the stylized facts for all major macroeconomic time series in Section 4. In Section 5 we reconcile data with the model with a unique minimum distance exercise. Section 6 follows with simulation exercises and we conclude in Section 7 with a re-statement of our central contribution: that given the recurrent manner in which macroeconomic time series exhibit large deviations, LRMN model representations may be more suitable to empirical analyses of linear DSGE models.

2. Rational Expectations vs. Adaptive Learning

Linear DSGE models begin by specifying a familiar form

$$X_t = A(\theta)E_t(X_{t+1}) + B(\theta)U_t, \tag{2}$$
$$U_t = PU_{t-1} + \varepsilon_t,\ E(\varepsilon_t\varepsilon_t') = \Sigma(\theta), \tag{3}$$

where θ denotes a vector of parameters, X_t denotes all model variables (usually in logarithmic deviations from respective steady state values) and ε_t denotes innovations to structural shock processes (U_t) with $\Sigma(\theta)$ denoting the variance-covariance matrix of the innovations. The next step in preparing a model for empirical analysis under RE is to replace expectations with realizations and idiosyncratic expectational errors (η_t), one then obtains the form employed by Sims (2001)

$$\Gamma_0(\theta)X_t = \Gamma_1(\theta)X_{t-1} + \Psi(\theta)\varepsilon_t + \Pi(\theta)\eta_t,\ E(\varepsilon_t\varepsilon_t') = \Sigma(\theta). \tag{4}$$

The RE solution to the above system yields the state equation that describes the evolution of all model variables

$$X_t = F(\theta)X_{t-1} + G(\theta)\varepsilon_t,\ E(\varepsilon_t\varepsilon_t') = \Sigma(\theta), \tag{5}$$

which, when coupled with an observer equation

$$Y_t = H'X_t, \tag{6}$$

where H is a selection matrix that links de-trended macroeconomic data (Y_t) to model variables (X_t), yields the state-space empirical representation of a linear DSGE model in (5)-(6) under RE. Specifying a Normal distribution for the innovations then allows one to use a Kalman Filter to write down the likelihood function ($L(Y_t|\theta)$) associated with a model employed in estimating the parameters using, for example, either Maximum Likelihood or Bayesian full-information techniques. Given parameter estimates, one can employ the state equation to compute the usual objects of inquiry (e.g. impulse responses, variance decompositions etc.).[4] The key to such an empirical exercise is data-model congruency: the assumption on the underlying data generating process (DGP) is that macroeconomic data

[4]DeJong and Dave (2011) describe the details of implementing limited or full information empirical exercises with either linear or non-linear DSGE models.

in their cyclic form follow a fixed-coefficient VAR

$$Y_t = \Theta Y_{t-1} + \upsilon_t, \;\; E(\upsilon_t \upsilon_t') = \Omega, \tag{7}$$

and the model also follows this form in that the state equation (5) takes the form of a fixed-coefficient VAR. This methodology, sometimes with extensions and variants (e.g. DSGE-VARs) then forms the core of what is known as RE macroeconometrics.

The assumption that the DGP for Y_t is a fixed-coefficient VAR is exactly that, an assumption on an unknown DGP. In particular this assumption implies that the assumed distribution for υ_t translates into a similar stationary distribution of Y_t, e.g. assuming $\upsilon_t \sim N(0, \Omega)$ implies that data are also Normal with thin tails. So, under this assumption, in order to generate large movements (volatility) in Y_t (e.g., rare but deep recessions or booms) one can assume a fat-tailed distribution for υ_t (and correspondingly for ε_t) so that the conclusion is 'fat tails in, fat tails out'. Alternatively one could introduce stochastic volatility directly into the specification of the stochastic properties of υ_t (and correspondingly for ε_t), or equivalently assume that rare disasters strike the above process directly (as in Reitz, 1988 or Barro, 1999). Finally one could argue, as we do, that a change in fundamental assumptions so that the DGP does not take the form of a fixed-coefficient VAR but instead follows a LRMN, is warranted.

Assuming that the DGP follows a LRMN amounts to assuming that instead of (7), data follow

$$Y_t = \Phi_t Y_{t-1} + u_t \tag{8}$$

where the noise terms u_t *and* Φ_t are stochastic processes. Given precise assumptions on the nature of stochasticity of the two processes, the applied mathematics and statistical theory literature (starting with Kesten, 1973), suggests that the tail of the stationary distribution of Y_t can be fat *despite* u_t following a thin-tailed distribution. If a model were to follow this form then the conclusion would be 'thin tails in, fat tails out'. This is because a LRMN accumulates fundamentally differently than a fixed-coefficient VAR: even if $E(\Phi_t) < 1$ it is still the case that $P(\Phi_t > 1) > 0$. This translates into the tail of the stationary distribution of Y_t being distributed Y^{-p} where $p - 1$ measures the *number* of moments that exist for the underlying distribution. When p is small, only $p - 1$ moments exist hence the tail is fatter while under Normality p is large and all moments exist. So this avenue of endogenously generating large movements in Y_t in response to thin tailed shocks is to assume that the underlying DGP follows a LRMN. How would an actual DSGE model then deliver the same form in response to a change in a fundamental economic assumption? Under RE the state equation describing model dynamics follows a fixed-coefficient VAR. We argue that changing assumptions on expectations formation, from RE to constant gain stochastic gradient (SGCG) adaptive learning, leads to a model representation being of the LRMN form.

Following a large literature described in Evans and Honkapohja (2001), under adaptive learning agents are presumed to only employ information up until time $t - 1$ in forming the required expectations. This alters the original specification of the model to

$$X_t = A(\theta) E_{t-1}(X_{t+1}) + B(\theta) U_t. \tag{9}$$

Under adaptive learning agents estimate their model of the dynamics of economic variables, called the perceived law of motion (PLM), by recursive least squares (RLS), arguably the most common estimation method in econometrics. The PLM has the same functional form as the RE equilibrium (REE) but possibly different coefficients (b_t) since agents do not know the REE. To estimate the PLM, agents use past data and then generate forecasts using the estimated model. Thus, a perceived law of motion (PLM) on the part of agents is conjectured as a time varying analog to the RE solution:

$$X_t = b_t X_{t-1} + \xi_t,\ \xi_t \sim iid(0, \Xi), \tag{10}$$

which implies, given the timing assumption, that

$$E_{t-1}(X_{t+1}) = b_{t-1}^2 X_{t-1}.$$

Inserting this expression into (9) yields an actual law of motion (ALM) as

$$X_t = A(\theta) b_{t-1}^2 X_{t-1} + B(\theta) U_t. \tag{11}$$

Finally, a specification for a learning rule that governs the evolution of b_t is required; assuming a SGCG rule yields

$$b_t = b_{t-1} + g X_{t-1}(X_t - X_{t-1}' b_{t-1}) \text{ (for a given } b_0). \tag{12}$$

Such a rule has been shown to be optimal by Evans et al. (2010). The next step is to insert the ALM in place of X_t in order to derive a LRMN for the coefficients (b_t). Large deviations in b_t would then drive the same in X_t. However this raises analytical issues since the driving process (U_t) and innovations (ε_t) are far too embedded in the model specification. That is, it is not possible to represent the equation for b_t solely as a function of the structural shocks U_t and/or innovations ε_t having inserted the ALM into the SGCG learning rule. Therefore we consider the following alternate route to demonstrate how under learning model dynamics follow a LRMN.

Usually the PLM is formed by assuming something close to a RE solution. For example, above we assume that the PLM is a time varying VAR(1) since the RE solution takes a VAR(1) form. Instead let's assume that the PLM is a time varying moving average (MA) process

$$X_t = \Phi_t U_t. \tag{13}$$

Next, assume that $E_{t-1}(X_{t+1}) = \Phi_{t-1} U_t$ which yields the ALM

$$X_t = A(\theta) \Phi_{t-1} U_t + B(\theta) U_t. \tag{14}$$

The evolution of Φ_t is governed by the learning rule

$$\Phi_t = \Phi_{t-1} + g U_{t-1}(X_t - \Phi_{t-1} U_{t-1}) \tag{15}$$

and inserting the ALM in place of X_t in the equation above yields

$$\Phi_t = [I + gA(\theta)U_{t-1}U_t - gU_{t-1}U_{t-1}]\,\Phi_{t-1} + gB(\theta)U_{t-1}U_t. \tag{16}$$

The learning model can therefore be written as

$$\begin{aligned}
\Phi_t &= \Lambda_t\Phi_{t-1} + \Omega_t, && (17)\\
\Lambda_t &= I + gA(\theta)U_{t-1}U_t - gU_{t-1}U_{t-1},\ g \in (0,1), && (18)\\
\Omega_t &= gB(\theta)U_{t-1}U_t, && (19)\\
U_t &= PU_{t-1} + \varepsilon_t. && (20)
\end{aligned}$$

Next we recognize that the VAR(1) process for structural errors can always be written in its Wold representation yielding the LRMN system

$$\begin{aligned}
X_t &= \Phi_t P(L)\varepsilon_t && (21)\\
\Phi_t &= \Lambda_t\Phi_{t-1} + \Omega_t, && (22)\\
\Lambda_t &= I + gA(\theta)P(L)^2 L\varepsilon_t\varepsilon_t - gP(L)^2L^2\varepsilon_t\varepsilon_t,\ g \in (0,1), && (23)\\
\Omega_t &= gB(\theta)P(L)^2L^2\varepsilon_t\varepsilon_t. && (24)
\end{aligned}$$

In this representation the multiplicative term Λ_t in equation (22) is now only a function of the innovations ε_t in equation (23) suggesting that (following Kesten, 1973) the tail of the stationary distribution of the coefficients Φ_t may follow a power law. This would then impart fat tails for X_t as well since $X_t = \Phi_t U_t$.[5] To show that the tail of the stationary distribution of model variables X_t is fat as the constant gain (g) increases, we map out the relevant relationship with simulations. The intuition behind the simulations we provide is straightforward: LRMN specifications suggest that small shocks (as is usually assumed in linear DSGE modeling) accumulate in a particular way so as to lead to large movements in model variables. Since the only assumption that has been altered is that of expectations formation (from RE to SGCG learning) and given that SGCG learning has a straightforward interpretation (that a high constant gain reflects either excessive structural change or agents have short memories), the conclusion is straightforward: given a high constant gain (g) it may be the case that history repeats itself due to short memories; macroeconomic variables exhibit predictably rare but large deviations from trend.

3. Related Literature

Four strands of the existing literature are related to our analyses of fat tails in macroeconomic data. First, recent empirical findings suggest that data on macroeconomic variables of interest such as output growth and inflation are seldom Normally distributed. In particular, the unconditional distribution of macroeconomic variables is often not Gaussian: Christiano (2007) provides pre-Great Recession evidence and Ascari et al. (2015) provides the same for

[5] The multivariate theory of recurrence relations with non *iid* shocks is not yet available (see Benhabib and Dave (2014) and references therein for a review of that literature).

more recent time periods. Further, Fagiolo et al. (2008) show that in the USA and many OECD countries, output growth-rate distributions can be well approximated by exponential power densities with tails much fatter than those of a Normal distribution, implying that output growth patterns tend to be quite lumpy. Large growth events, either positive or negative, seem to be more frequent than what a Gaussian assumption would predict.

The second strand of literature addresses the above empirical regularities in macroeconomic models. Broadly speaking, fat tails for endogenous variables in DSGE models emerge via two basic sources of non-Normality. The first source replaces the Normality of shocks that hit the economy with fat tailed shocks such as those distributed Laplace (see for e.g., Ascari et. al, 2015) or Student's-t (see for e.g. Curdia et. al., 2014 and Chib and Ramamurthy, 2014). In particular, Ascari et. al. (2015) show that exogenous fat tailed shocks in workhorse models do not always translate into corresponding extreme events in macroeconomic time series; they suggest a need for endogenous mechanisms that deliver such movements. Curdia et. al. (2014), on the other hand, use Smets and Wouter's (2007) model to show that the Gaussianity assumption in DSGE models is questionable, even after allowing for low-frequency changes in the volatility of shocks. Moreover, their analyses show that a Student's-t specification is strongly preferred by the data. However, the fat tails delivered under these models could directly be the result of exogenous stochastic disturbances whose distribution is non-Gaussian. Moreover, simply replacing Normally distributed shocks with non-Gaussian shocks is not without caveats. In particular, Müller (2013) describes some of the dangers associated with departures from Gaussianity when the alternative shock distribution is also misspecified. In contrast, the second source attributes fat tails in macroeconomic variables due to structural assumptions, even if a model is hit by purely thin-tailed uncorrelated innovations.

In this second channel, a natural point of departure from the traditional DSGE models is to consider an alternative DGP, for example time-varying parameter specifications for VARs, as follows. As a result, in the third strand of the literature, with respect to reduced form analyses, Cogley and Sargent (2001, 2005) suggest that attributing adaptive behavior to agents produces a macroeconomic model whose tendencies toward a self-confirming equilibrium are interrupted by recurrent escapes. Such tendencies generate nonlinearities in the data that can show up as drifting coefficients. Hence they specify time varying VARs in addition to assuming stochastic volatility in attendant innovations. Monache and Paterella (2014) build on the framework of Cogley and Sargent (2001) by replacing their Normally distributed shocks with Student's-t shocks, their modification finds support in the data.[6] What these papers suggest is that underlying adaptive algorithms can help match features of the data including possibly non-Normal characteristics.

Thus, in the fourth strand of the literature, Milani (2011) evaluates the empirical role of expectational shocks on business cycle fluctuations and relaxes the RE assumption to exploit survey data on expectations in the estimation of a NK model under learning. Milani (2005) studies whether learning can provide a reasonable source of the observed persistence in inflation and find that persistence heavily depends on the assumed learning speed. Primiceri (2009) further contributes to the discussion by demonstrating that in models exhibiting

[6] In contrast, Sims (1980, 1999) and Bernanke and Mihov (1998a, 1998b) support the time invariant view of macroeconomic data.

self confirming equilibria as in Cho et al. (2002) "...prolonged episodes of high inflation ending with rapid disinflations can occur when policymakers underestimate [in a learning algorithm] both the natural rate of unemployment and the persistence of inflation in the Phillips curve.". This reflects a sort of fat-tailed behavior for inflation as a function of constant gain learning. Marcet and Nicolini (2005) also consider a learning mechanism that produces small departures from RE within the model to match episodes of hyperinflation.

A related literature also considers the role of non-linearity originating from state dependence in DSGE models to understand the dynamics of the economy (e.g., Auerbach and Gorodnichenko, 2012, Ferraresi et al., 2015, Canzoneri et al., 2016). The non-linearity is incorporated in DSGE models by the possibility of a regime switch, for e.g., Franta (2015) introduced multi-regimes in the shock propagation mechanism while in the context of learning, Marcet and Nicolini (2005), Branch and Evans (2007) and Milani (2014) allow for a time-varying learning speed (regime switching) to generate endogenous volatility in output and inflation. Massaro (2013) instead exogenously assumes multiple heuristics where agents discretely shift between rules for guiding expectation formation in an adaptive learning model.

Similar to the reduced form and adaptive learning macroeconometric exercises discussed above, we also appeal to the notion that the underlying DGP of macroeconomic variables might not follow fixed-coefficient VAR specifications. However our point of departure is not time or state varying specifications but instead the possibility that data follow LRMN processes. The key difference is that the multiplicative *and* additive terms in such specifications are themselves stochastic processes. In the applied statistical theory and mathematics literatures such specifications have been analyzed in order to determine the limiting properties of these systems. For example, Kesten (1973) and Goldie (1991) examine the multivariate case of a LRMN when the multiplicative term follows an *iid* process, Saporta (2005) examines the case when that same term follows a Markov process and Roitershtein (2007) builds on that literature by looking at multiplicative terms that are Markov modulated. In this literature the key is the fact that the variable(s) of interest being modeled as LRMNs have stationary distributions whose tail is approximated by a power law. The smaller the coefficient on that power law, the fatter the tail. These technical results have been extensively employed in the macroeconomic literature: Benhabib (2009) examines regime switching monetary policies and characterizes the moments of the resulting stationary distribution of inflation, Benhabib et al. (2011) model the evolution of wealth in an OLG model with stochastic capital and labor income and Benhabib and Dave (2014) examine the evolution of the price-dividend ratio in US data using a univariate LRMN that results from the assumption of a constant gain adaptive learning algorithm.

4. Power Law Stylized Facts

Clauset et al. (2009) provide a statistical framework for quantifying power law behavior in data and a procedure for measuring the fatness of tails. In this section, we first briefly explain Clauset et al.'s (2009) approach which we apply to our macroeconomic data. We then describe our data, present the associated fat tails measures and finally discuss the interpretation of our findings.

4.1. Tail Index Estimation: Method

Clauset et al.'s (2009) statistical framework seeks to quantify an interpretable index for the thickness of the tails of a time series. In particular, a time series obeys a power law if drawn from a probability distribution ($P(Y)$) of the form

$$P(Y) \propto Y^{-p}, \tag{25}$$

where p is a constant parameter of the distribution-known as the exponent, scaling or tail index. Only the values above some minimum threshold (denoted as $Y_{\min}$) follow a power law distribution and in such cases we say that the tail of the distribution is thick. Estimation of that index proceeds as follows. Taking logarithms of equation (25) yields

$$\ln(P(Y)) = -p\ln(Y) + C, \tag{26}$$

where C is a constant and $\ln(P(Y))$ captures the frequency of data, denoted by Y. Constructing a histogram representing the frequency distribution of observed Y and plotting that histogram on a log-log plot implies that a power law follows a straight line with a slope of $-p$. This slope determines the thickness of the tail and can be estimated via a least squares linear regression. A steeper negative slope with a higher p implies that the probability of larger values of $\ln(Y)$ are less frequent compared to a flatter negative slope with a smaller p. This procedure for estimating the tail index dates back to Pareto's work on wealth distributions (see Arnold, 1983). However, Clauset et al. (2009) show that such a least squares methodology is subjective in estimating the slope, and may lead to systematic and significant errors. Though the underlying intuition behind the estimation of tail index is the same, Clauset et al. (2009) provide a maximum likelihood (ML) estimator, where under the assumption that data are drawn from a distribution that follows a power law exactly for some $Y_{\min} \leq Y$, a likelihood function yields an estimate of the form

$$\widehat{p} = 1 + n\left[\sum_{i=1}^{n} \ln\frac{Y_i}{Y_{\min}}\right]^{-1}. \tag{27}$$

where $\widehat{p}$ denotes the estimate derived from data Y_i, $i = 1, ..n$, and $Y_{\min} \leq Y_i$.[7] The width of the likelihood maximum estimator provides standard errors for $\widehat{p}$, given by

$$\sigma = \frac{\widehat{p}-1}{\sqrt{n}} o\left(\frac{1}{n}\right). \tag{28}$$

For details of derivation of the likelihood function and the ML estimator see Appendix B of Clauset et al. (2009). As in a simple log-log case, the interpretation of $\widehat{p}$ as the tail index is similar. Additionally $\widehat{p}-1$ measures the number of moments that exist for the underlying

[7]Note that Clauset et al.'s (2009) method estimates Y_{min} and p according to the goodness-of-fit based method described in Clauset, Shalizi, Newman (2007). The fitting procedure works as follows: 1) For each possible choice of Y_{min}, p is estimated via the method of maximum likelihood, and the Kolmogorov-Smirnov goodness-of-fit statistic (denoted by D) is calculated. 2) Value of Y_{min}, that gives the minimum value D over all values of Y_{min}.is then selected as the estimate of Y_{min}.

distribution. If $\widehat{p}$ is small then the likelihood of $Y_i \geq Y_{\min}$ is high, higher moments of the distribution are infinite and therefore an extreme event is more frequent hence the tail is fatter, and vice versa.

Apart from power laws, several distributions can exhibit fat tails, for example the exponential (EP) distribution. The main difference between the EP versus power law distributions is that the EP density is characterized by exponentially shaped tails which are thicker than that of the Normal distributions but thinner than that of the power law distributions. Moreover, unlike power law distributions, EP (albeit tails fatter than Normal distributions) are characterized by finite moments of any order. However, empirical evidence of the presence of structural breaks in the mean growth rate of the USA and of other OECD countries (Stock and Watson, 1999, Ben-David et al. 2003) and the recent debate on the 'Great Moderation' (i.e., the decline in output volatility observed since the end of the 1980s as documented by McConnell and Perez-Quiros, 2000, Blanchard and Simon, 2001, Stock and Watson, 2002, Kim et al., 2004) suggest that higher moments of GDP growth rates may not be stable or finite. The GDP growth rate therefore is more likely to follow a power law distribution (with smaller $\widehat{p}$ such that the higher moments are infinite) than the EP (with fatter tails but finite moments of any order).[8]

In the next sub-section we use this procedure to estimate the tail index of our data and present the results.

4.2. Data

The FRED database provides the raw series on output, consumption, investment, prices, population, money stocks and interest rates (GDPC96, PCECC96, GPDIC1, GDPDEF, CNP16OV, M2SL and TB3MS respectively). Note that the last three time series are monthly so we compute monthly averages to obtain data at a quarterly frequency. With the raw data in hand, per-capita series are constructed as:

$$Y_t = \frac{(GDPC96_t/4)}{CNP16OV_t} \times 1000000, \; C_t = \frac{(PCECC96_t/4)}{CNP16OV_t} \times 1000000, \quad (29)$$

$$I_t = \frac{(GPDIC1_t/4)}{CNP16OV_t} \times 1000000, \; M_t = \frac{(M2SL_t/GDPDEF_t) \times 100}{CNP16OV_t} \times 1000000, \quad (30)$$

$$P_t = \frac{GDPDEF_t}{GDPDEF_{t-1}}, \; R_t = \frac{1}{(1 - TB3MS_t/400)}. \quad (31)$$

Next we denote the natural logarithms of the above variables as $y_t = \log(Y_t)$, $c_t = \log(C_t)$, $i_t = \log(I_t)$, $m_t = \log(M_t)$, $\pi_t = \log(P_t)$ and $r_t = \log(R_t)$. Each element of this set of time series $\{y_t, c_t, i_t, m_t, \pi_t, r_t\}$ can then be detrended using the various standard methods described in DeJong and Dave (2011) in order to generate cyclic series, denoted as $\{\widehat{y}_t, \widehat{c}_t, \widehat{i}_t, \widehat{m}_t, \widehat{\pi}_t, \widehat{r}_t\}$.

[8] Note that it is difficult to directly compare the estimates of the tail index of the detrended data under a power law distribution from the current paper with the EP of raw data (e.g. Fagiolo, 2008).

4.3. Tail Index Estimation: Results

For each of the series described above we report: estimates of its tail index ($\widehat{p}$) along with associated standard errors ($se(\widehat{p})$) in Table 1 and estimates of kurtosis in Table 2. Since Canova (1998) showed that the choice of detrending method can affect the stylized facts of business cycle, we estimate the tail indices under all of the various detrending methods usually employed in DSGE models.

Table 1. Tail Index Estimates For Cyclical Components

	First Difference		Linear Trend		HP-Filter		CF-Filter		BK-Filter	
	$\widehat{p}$	$s.e.(\widehat{p})$	$\widehat{p}$	$s.e.(\widehat{p})$	$\widehat{p}$	$s.e.(\widehat{p})$	$\widehat{p}$	$s.e.(\widehat{p})$	$\widehat{p}$	$s.e.(\widehat{p})$
$\widehat{y}_t$	5.1386	3.6723	3.7068	8.9334	3.6395	0.6942	3.3663	0.4739	4.2914	1.8918
$\widehat{c}_t$	6.1485	2.6279	10.9348	13.3347	3.2730	1.3030	3.3482	0.6886	5.4777	1.7408
$\widehat{i}_t$	18.0209	6.1364	6.0663	2.3156	2.8654	2.3411	3.9022	0.8682	5.8127	8.0510
$\widehat{m}_t$	3.3886	0.8516	2.2667	0.3808	5.9994	2.1795	3.4925	1.1075	8.5404	3.3866
$\widehat{\pi}_t$	4.2283	1.4194	5.4529	2.0732	3.0590	0.5232	2.5105	0.2821	2.6515	0.2879
$\widehat{r}_t$	3.1239	1.0418	3.2306	0.7263	3.9196	1.9271	4.1979	2.5761	4.9127	1.6802

Table 2. Kurtosis of Cyclical Components

	First Difference	Linear Trend	HP-Filter	CF-Filter	BK-Filter
$\widehat{y}_t$	4.2901	2.1189	3.3810	3.8256	3.3988
$\widehat{c}_t$	5.4227	1.9943	3.2046	3.9425	3.0538
$\widehat{i}_t$	5.1690	2.6129	4.3114	4.0372	3.8558
$\widehat{m}_t$	5.5172	3.9285	3.0036	3.7455	2.3400
$\widehat{\pi}_t$	4.2940	4.4868	6.6516	7.9359	6.5073
$\widehat{r}_t$	14.5214	4.1277	3.6099	2.9399	2.9881

Qualitatively, Table 1 shows that regardless of the detrending methods, the tail indices of all the major variables used in structural empirical macroeconomics are small pointing towards fatter tails. Table 2 shows that the same conclusion about the fat tails of detrended data can also be attained by using kurtosis which is in most cases bigger than 3 (the value for a Normal distribution), again pointing towards fatter tails. Both indices provide a strong indication that the assumption of Gaussianity may not hold in the data, and there is strong evidence of fat tails for all the real and nominal variables used in DSGE models. In the empirical implementation and simulation exercises reported below we focus on the tail indices given the relationship to LRMN model representations discussed above. Recall, if a series has an index of $\widehat{p}$ then, as discussed previously, the tail of the stationary distribution of that time series only has $\widehat{p}-1$ moments - e.g. if the $\widehat{p}$ for inflation is 3 then only the mean and variance of inflation may exist as moments. Given congruency with LRMN model representations we therefore employ these $\widehat{p}$ as our empirical targets and stress that while $\widehat{p}$ itself does not represent a moment, it does suggest *how many* moments a variable may have.

Quantitatively, we do observe some differences across the estimates and the associated standard errors of the tail indices, for example under linear and first-difference detrending, the estimates differ from the estimates of other detrending methods and the standard errors are bigger. However, both the estimates and the standard errors of the tail indices for other filtering techniques are quite comparable. In the empirical implementation and simulation exercises reported below we focus on the tail indices estimated for HP-filtered data although other techniques should yield similar results. We use the HP-filter $\widehat{p}$ estimates because the estimates are significant and HP filtered data is most commonly employed in the DSGE models.

5. Structural Estimation

5.1. Method

We denote a column vector of empirical targets (e.g. tail indices from Table 1) as $\varkappa$ and note that coupling the time varying VAR(1) form of a model's solution with an observer equation (using HP filtered data) allows us to calculate via a Kalman smoother the smoothed values of the state vector. Thus for a given parameterization of μ the VAR(1) model representation delivers a smoothed series of the same length as the data. The smoothed series then allow a calculation of corresponding model targets given a parameterization for μ. The tail indices for these smoothed series can also be estimated using the methods of Clauset et al. (2009) and we denote them in a column vector as $\varkappa(\mu)$. We next search over the parameter space to minimize the squared difference between $\varkappa$ and $\varkappa(\mu)$ in order to estimate values for μ; that is, our estimates are delivered by

$$\min_{\mu} F = [\varkappa - \varkappa(\mu)]'[\varkappa - \varkappa(\mu)] \tag{32}$$

with standard errors computed using the Hessian of the above objective function at the parameter estimates. Our main claim is that this minimum distance estimation method is not just distribution free but also does not necessarily entail the matching of any particular set of moments if the empirical targets are not moments but tail indices. Why is our method distribution free? We note that nowhere in the model specifications have we assumed that the *iid* structural shocks follow any particular distribution. Further, we note that the use of a Kalman smoother does not require the assumption of a distribution for errors in a state-space; it is a calculation exercise and a Kalman filter is only optimal in the event that additive noise is assumed to be Gaussian, see Lutkepohl (2007) and references therein.[9] We do not need for the filter nor the associated likelihood function to be optimal since we do not use those objects; all we need is that given a parametrization for μ that the Kalman smoother allows us to calculate the smoothed values from a state space specification.

Our method does not *necessarily* entail the matching of any moments if the empirical targets ($\varkappa$) are only tail indices. Recall from the discussion above that the tail index associated with a LRMN specification specifies the *number* of finite moments associated with a time series, and not any particular set of moments. We next describe the particular DSGE model

[9] Following Lutkepohl (2007) we note that "...it is possible to justify the Kalman filter recursions even if the initial state and the white noise processes are not Gaussian. In that case, the quantities obtained by the recursions are no longer moments of conditional normal distributions, however. For other interpretations of the quantities see, for example, Schneider (1988)."

we implement and then present estimation results. Step by step details on the estimation procedure are provided in an Appendix: Estimation Procedure.

5.2. Results

For our structural model we adopt a three equation New-Keynesian Model (NKM) framework in which a system is written for deviations of output from trend (y_t), inflation (π_t) and nominal interest rates (r_t) with three structural processes (ζ_{1t}, ζ_{2t} and ζ_{3t}) that respectively reflect preference changes in the Euler equation, exogenous changes in the marginal costs of production in the Phillips curve and a shock to policy. The system is given by

$$\begin{aligned}
y_t &= E_t(y_{t+1}) - \tau r_t + \tau E_t(\pi_{t+1}) + \zeta_{1t}, && (33)\\
\pi_t &= \beta E_t(\pi_{t+1}) + \kappa y_t + \zeta_{2t}, && (34)\\
r_t &= \theta r_{t-1} + (1-\theta)\gamma_1 \pi_{t-1} + (1-\theta)\gamma_2 y_{t-1} + \zeta_{3t}, && (35)\\
\zeta_{1t} &= \rho_1 \zeta_{1t-1} + \varepsilon_{1t},\ \varepsilon_{1t} \sim iid(0, \sigma_1^2), && (36)\\
\zeta_{2t} &= \rho_2 \zeta_{2t-1} + \varepsilon_{2t},\ \varepsilon_{2t} \sim iid(0, \sigma_2^2), && (37)\\
\zeta_{3t} &= \rho_3 \zeta_{3t-1} + \varepsilon_{3t},\ \varepsilon_{3t} \sim iid(0, \sigma_3^2). && (38)
\end{aligned}$$

This model takes the form of a time varying state equation when we include SGCG learning in this framework

$$\begin{aligned}
X_t &= \Gamma b_{t-1}^2 X_{t-1} + \Psi U_t && (39)\\
b_t &= b_{t-1} + g X_{t-1}(X_t - X_{t-1}' b_{t-1}),\ b_0 \text{ given.} && (40)
\end{aligned}$$

to which we append an observer equation of the form

$$Y_t = H' X_t. \tag{41}$$

We use the state space above and a parameterization for μ to obtain smoothed values for our variables of interest using a Kalman smoother. Since we have three endogenous model variables, we need to employ data for two of them in order to obtain smoothed values of the third variable from the model using the smoother. As a result, we have model smoothed series for output, inflation and interest rate, given a parameterization for μ. For each of the smoothed series on output, inflation and interest rate we estimate its respective tail index and minimize the distance from the tail index estimated from the actual data by choice of parameters (including g).

Our first estimation results minimize the objective function defined as the distance between the tail indices for all the variables jointly from the filtered data and the model based Kalman smoothed data, as follows:

$$\min_{\mu} F = \left[\begin{pmatrix} p^y \\ p^\pi \\ p^r \end{pmatrix} - \begin{pmatrix} \widehat{p}^y(\mu) \\ \widehat{p}^\pi(\mu) \\ \widehat{p}^r(\mu) \end{pmatrix} \right]' \left[\begin{pmatrix} p^y \\ p^\pi \\ p^r \end{pmatrix} - \begin{pmatrix} \widehat{p}^y(\mu) \\ \widehat{p}^\pi(\mu) \\ \widehat{p}^r(\mu) \end{pmatrix} \right], \tag{42}$$

where p^i, $i = y, \pi, r$ denotes the tail index of a series and $\widehat{p}^i(\mu)$, $i = y, \pi, r$ denotes the tail index from the corresponding smoothed series, given a parametrization μ.

Table 3 presents the main results for this estimation exercise. In column 2 we present the estimates from Milani (2011) as a reference point and column 3 presents parameter estimates (and respective standard errors).

Table 3. Matching Tail Coefficients

		All Series
	Milani	Est. (Std. Err.)
γ_1	1.4170	1.4093 (0.0001)
γ_2	0.2210	0.2202 (0.0001)
θ	0.9498	0.9513 (0.0001)
κ	0.0350	0.0345 (0.0004)
τ	0.2360	0.2317 (0.0004)
ρ_1	0.3538	0.3605 (0.0001)
ρ_2	0.1746	0.1791 (0.0062)
ρ_3	N/A	0.1821 (0.0434)
β	0.9615	0.9628 (0.0012)
g	0.0196	**0.0513 (**0.0001)
σ_1	0.7700	0.7633 (0.0110)
σ_2	0.2970	0.2903 (0.0035)
σ_3	0.2070	0.2055 (0.0410)
F		3.0081E-13

We next re-estimate the parameters by minimizing an objective function for each of the variables separately. The corresponding objective functions are as follows:

$$\min_{\mu} F = [p^y - \widehat{p}^y(\mu)]' [p^y - \widehat{p}^y(\mu)] \tag{43}$$

$$\min_{\mu} F = [p^\pi - \widehat{p}^\pi(\mu)]' [p^\pi - \widehat{p}^\pi(\mu)] \tag{44}$$

$$\min_{\mu} F = [p^r - \widehat{p}^r(\mu)]' [p^r - \widehat{p}^r(\mu)] \tag{45}$$

Table 4. Matching Tail Coefficients

	Milani	y_t Est. (Std. Err.)	π_t Est. (Std. Err.)	r_t Est. (Std. Err.)
γ_1	1.4170	1.4148 (0.7186)	1.4115 (0.0001)	1.4195 (0.0001)
γ_2	0.2210	0.2221 (0.1003)	0.2199 (0.0001)	0.2239 (0.0001)
θ	0.9498	0.9483 (0.4206)	0.9498 (0.0001)	0.9513 (0.0001)
κ	0.0350	0.0354 (0.1537)	0.0352 (0.0001)	0.0353 (0.0006)
τ	0.2360	0.2426 (0.7169)	0.2355 (0.0009)	0.2387 (0.0001)
ρ_1	0.3538	0.3551 (0.0001)	0.3577 (0.0017)	0.3589 (0.0001)
ρ_2	0.1746	0.1819 (0.0001)	0.1761 (0.0002)	0.1759 (0.0001)
ρ_3	N/A	0.1797 (0.0001)	**0.4411** (0.0001)	0.1758 (0.0001)
β	0.9615	0.9624 (0.2153)	0.9616 (0.0039)	0.9624 (0.0001)
g	0.0196	**0.0515** (0.0057)	**0.0685** (0.0001)	**0.0693** (0.0001)
σ_1	0.7700	0.7744 (0.0001)	0.7747 (0.0001)	0.7751 (0.0001)
σ_2	0.2970	0.2814 (0.0001)	0.2998 (0.0042)	0.2842 (0.0001)
σ_3	0.2070	0.2041 (0.0001)	0.2080 (0.0011)	0.2080 (0.0001)
F		8.6674E-12	1.7201E-11	3.6299E-12

Table 4 presents the main results for this second estimation exercise: columns 3, 4 and 5 respectively present parameter estimates (and standard errors) for output, inflation and the interest rate. It must be noted that the estimated parameters apart from g are very close to the estimates used in the literature, which were also our starting values and therefore the standard errors corresponding to our estimates are not too large in Tables 3 and 4. Our main result from Tables 3 and 4 is that the estimated constant learning gain parameter (g) is much bigger than the estimate usually documented in the existing literature (between 0.01-0.03). In comparison, our estimated g is almost double in both estimation exercises. Intuitively, g is inversely proportional to the amount of past data utilized by the agents in forecasting future macroeconomic variables. Roughly speaking, an estimated value of 0.1 of g translates into approximately 35 years of time series data utilized by the economic agents as opposed to 60+ years of time series data which corresponds to the estimated value of 0.03 of g.

Given our focus on a larger than usual estimated value for the gain parameter, we next explore two main concerns. The first is whether our particular LRMN does indeed return low tail indices as the gain (g) increases for variables of interest. This is an important (simulated) comparative static to conduct since constant gains closer to zero are associated with rational expectations and thus model variables should have larger tail indices indicating proximity to Normality. Our second concern is the extent to which our estimates of the constant gain are identified from our minimum distance empirical implementation.

Recall that the LRMN representation of our system of interest is given by

$$
\begin{aligned}
X_t &= \Phi_t U_t,\ t = 1, ..., T && (46)\\
\Phi_t &= \Lambda_t \Phi_{t-1} + \Omega_t, && (47)\\
\Lambda_t &= I + g\Gamma U_{t-1} U_t - g U_{t-1} U_{t-1},\ g \in (0,1), && (48)\\
\Omega_t &= g\Psi U_{t-1} U_t, && (49)\\
U_t &= P U_{t-1} + \varepsilon_t,\ \varepsilon_t \sim iid(0, \Sigma). && (50)
\end{aligned}
$$

Our first step in conducting the simulations is to draw shocks (ε_t) from a known distribution. We choose a uniform distribution with support from 0 to 1 and draw 5000 series of length 1000. For each ε_t series of length $T = 1000$, we simulate the above system using as a baseline the parameter estimates from Table 4. This allows us to obtain 5000 simulated series for X_t for a given value of the gain parameter. For each of the 5000 series we estimate the tail index using the method of Clauset et al. (2009) and then average that estimate across the 5000 series. Thus, for a given gain parameter we obtain the average $\widehat{p}$. We then repeat the same process by varying g from 0 to 0.1. We plot our comparative statics for the estimated average $\widehat{p}$ corresponding to the constant gain learning parameter in Figures 1,2 and 3 below. The figures show that for all the variables of interest (X_t), the average tail index monotonically decreases as the value of the constant gain parameter increases. In essence, the model does seem to suggest fatter tails for the stationary distribution of simulated macroeconomic aggregates as the learning horizon decreases.

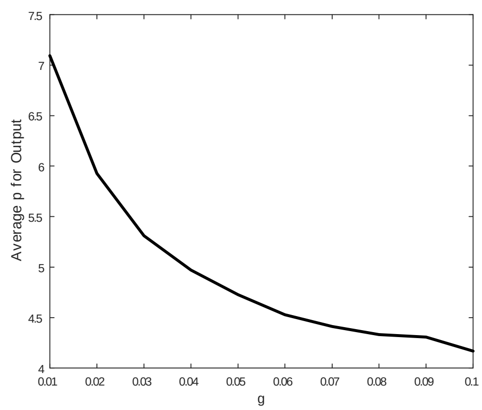


Figure 1. Simulated Output Tail Indices vs. g.

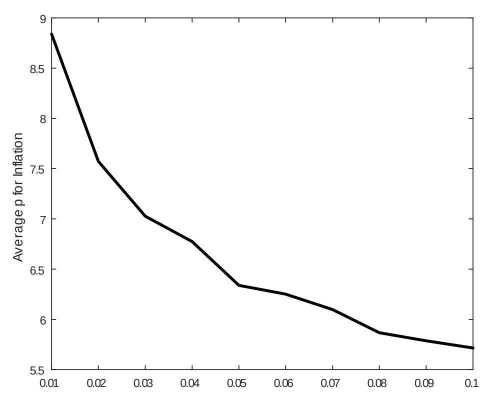


Figure 2. Simulated Inflation Tail Indices vs. g.

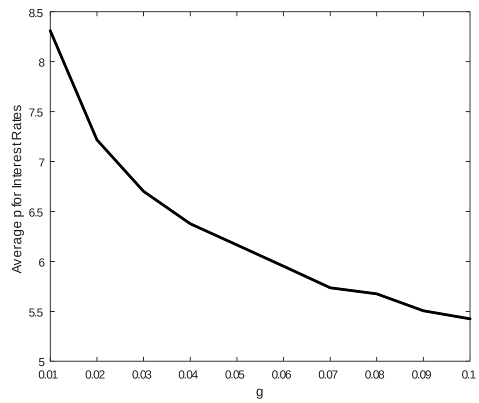


Figure 3. Simulated Interest Rate Tail Indices vs. g.

In any DSGE empirical implementation, the identification of parameters from data is a concern. The literature offers various approaches to estimate the parameters of a DSGE model: maximum likelihood, the method of moment, indirect inference etc (see DeJong and Dave, 2011). A common feature across these methods is that the underlying generic DSGE model is usually a linearized version with rational expectations and Gaussian shocks. Irrespective of estimation technique, DSGE models are constantly confronted with the issue of identification. Among recent contributions to addressing these identification issues in DSGE models are Iskrev (2008) and Iskrev (2010), Canova and Sala (2009), and Komunjer and Ng (2011). However, the issue of identification persists as noted by Canova and Sala (2009) for rational expectations environments and the issue may also affect models that depart from rational expectation and adopt specifications such as adaptive learning (Milani, 2012).

In our context we introduce adaptive learning in an otherwise standard NKM DSGE model. However, our model characterization alters the assumption on the underlying data generating process from a fixed coefficient VAR to a LRMN. This deviation has its trade-offs. While we aim to contribute to the literature in understanding the potential channels to incorporate the empirical regularity of fatter tails evident in data into a DSGE framework, a formal test for identification (as in Koumunjer and Ng, 2011) is beyond the scope of the current analysis. We therefore adopt various analyses to tackle potential identification issue in the last estimation exercise.

In particular, in our last empirical exercise, we estimated all free parameters in μ while minimizing the distance between the model and empirical tail indices. Given the possibility that an alternative value of parameters may minimize the value function, we perform two additional analyses to investigate our concerns. In our first analysis, we study how the value of the objective function in the last estimation changes for various values of g while keeping all the other parameters fixed at the estimated values. The illustrative analysis we undertake is a naïve way to identify any potential identification issues that may be of a concern if various values of g can deliver convergence of value function to numerical zero. We present this illustration in Figure 4.

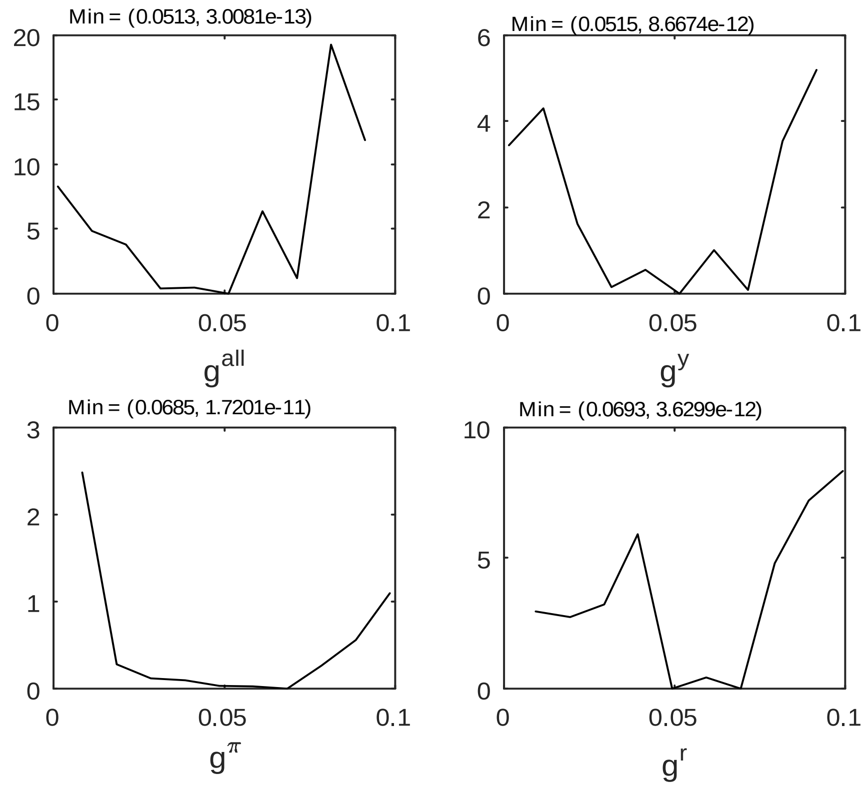


Figure 4. The Objective Function (F: Equation 42 - 45) vs. g.

It is clear that the minimum for the value function is only achieved at a certain estimate of g as provided in Tables 3 and 4. The last statement holds true for all versions of the objective functions except for the interest rate. For all the other objective functions, it is clear from the graphs that the value function approaches zero for other values of g but never converges to zero except at the unique value of g. For the interest rate, the value function converges to zero for multiple values of g. It is important to note that the graphical illustration is adjusted in order to avoid large values of the objective function especially when g is greater than 0.1. Despite this adjustment, some graphical illustrations can be deceptive to the naked eye and therefore we caution the reader that apart from the interest rate, for all other objective functions only a unique value of g and the corresponding converged value of the objective function mentioned at the top of each graph is the only solution. Based on the results from Tables 3 and 4 and the illustrative evidence from Figure 4, one can conclude that a large g is successful in delivering fat tails, so that a minimal role may remain for alternate structural parameters that govern transmission mechanisms and shock sizes or the persistence of shocks, to deliver fat tails.

In our second analysis of the identification issue, except for g we fix all the values of the parameters in μ at the values given in Table 3, column 3. We choose these values for two reasons. First, the estimates are not very different from Milani (2011) and second, our model also includes a persistence parameter (ρ_3) for the shock to policy which is not estimated by Milani (2011). We then estimate the value of g for all our objective functions presented in the last section. This analysis allows us to address the issue of estimating more parameters while minimizing the distance between the empirical and theoretical targets in the last section. We present the results of the estimated g and the corresponding standard errors in Table 5.

Table 5. Matching Tail Coefficients

	All series	y_t	π_t	r_t
	Est. (Std. Err.)	Est. (Std. Err.)	Est. (Std. Err.)	Est. (Std. Err.)
g	0.0513 (0.0086)	0.0518 (0.0090)	0.0514 (0.0086)	0.0313 (0.0014)
$\mathcal{F}$	6.2493E-4	1.5132E-4	2.1359E-14	1.4980E-17

As in Figure 4, we vary the value of g and investigate if the value of g where the objective function converges to numerical zero is unique. We illustrate these results in Figure 5. We urge the reader to read the illustrative evidence in light of the mentioned minimum value printed on top of each graph since graphs can be deceptive to naked eye. The main result from Table 5. and the illustration in Figure 5. is that the estimated g is close to the estimates from Table 3 and 4 and the estimates continue to be bigger than the estimate of g documented in the literature. As a result, we can conclude that while estimating g alone, there is a clear evidence that the estimated g is indeed unique and bigger than the estimated g documented in the past literature for all our objective functions barring the one associated with the interest rate. To elaborate further, the model provided above is suitable in studying the output and inflation dynamics, but not necessarily for the interest rate.

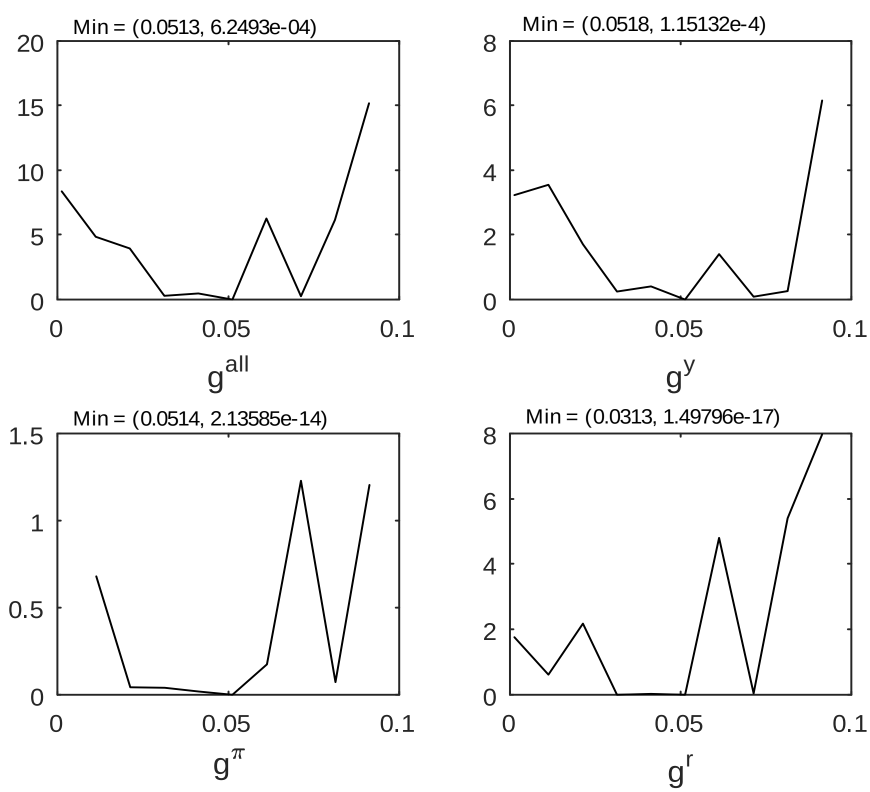


Figure 5. The Objective Function ($\mathcal{F}$: Equation 42 - 45) vs. g.

In summary, we see that in the baseline estimation and the estimation that accounts for potential identification issues find that a higher constant gain parameter g is successful in delivering fat tails. We further note that the estimates of the standard deviation of shocks either don't increase dramatically in the baseline estimation or our main result sustain when the shock size is fixed at the values as used in the literature, therefore we show that a larger than usual shocks are not driving the dynamics to deliver fat tails in the data under the LRMN assumption. We show in the next section, that a model without the constant learning mechanism is unable to match the statistical regularity of fatter tails in macroeconomic data. This evidence reassures us that absent learning mechanisms, other transmission mechanisms may not be enough to account for the fatter tails observed in data. In summary, we conclude that our core result of a higher constant gain is robust to all the estimation procedures provided in the prior sections.

It must be noted that the objective function which aims to minimize the distance between tail indices of the empirical and model based Kalman smoothed data are unsuccessful in minimizing the distance for the interest rate; for the estimates reported above for the interest rate column the objective function value F was not near machine zero. One potential reason is that we only use a reduced form policy equation for the interest rate and it may be the case that a more sophisticated model of the interest rate is needed. The estimate of g in this case is smaller than what we have reported earlier. However, we must be cautious in interpreting the estimates for interest rate in these cases. A smaller g may be successful in reducing the distance of the objective function but it is unable to deliver a good fit of the model as the function value is not close to zero (hence not minimized in our interpretation).

In the appendix, we also provide additional analyses on matching other empirical targets (i.e., the available moments) based on the information loaded in our tail indices. We present the corresponding tables and the illustrative evidence in the Appendix as well. The main conclusion that a DSGE with SGCG learning provides an endogenous channel for accounting for fatter tails is echoed from additional analyses as well.

In the next section, we provide evidence based on simulations from our theoretical model and its comparison with other alternatives in matching the empirical targets of tail indices.

6. Simulations

In this section, we assess whether our DSGE model under SGCG learning outperforms other plausible alternatives in accommodating fat tails as evidenced in the data. To do so, we simulate data under various models. For each of the model specifications, we draw 1000 series of ε_t of length greater than the empirical data from a specific distribution. We discard the first 100 and last 100 of the draws and only retain a series of the exact same length as in the empirical data. For each of the 1000 series, we simulate the model for y_t, π_t and r_t and estimate the respective fat tail index (p) using Clauset et. al's (2009) procedure. In particular, our first set of data is simulated using our DSGE model under SGCG learning and we call it the "Adaptive Learning Model". Under this model, ε_t is drawn from an *iid* Normal distribution with mean 0 and variance 1. Our second set of data is simulated by using exactly the same draws of ε_t as in the Adaptive Learning Model, but instead we close the model under RE. We call this specification the "Rational Expectations Model". Lastly, for our third simulated dataset, we again close the model under RE but instead replace the Normal distribution of ε_t with a non-Gaussian process that has a priori fatter tails. More specifically the ε_t for this alternative are drawn from a Student's-t distribution with a degree of freedom of 3 and 15. The degree of freedom of 3 embodies the idea that the world is quite far from Gaussian, and quite extreme. Like Curdia et al. (2014) we also choose the degree of freedom of 15 which captures the view that the world is not quite Gaussian, but not too far from Gaussianity either. We call these specifications the "Rational Expectations (t-dof3) Model" and "Rational Expectations (t-dof15) Model", respectively. For the simulated series from Adaptive Learning, Rational Expectations and Rational Expectations (t-dof15) models, we estimate the p's and illustrate the results in Figure 6. and Figure 7. Moreover, for the simulated series from Adaptive Learning, Rational Expectations and Rational Expectations (t-dof3) models, we estimate the p's and illustrate the results in Figure 8. and Figure 9.

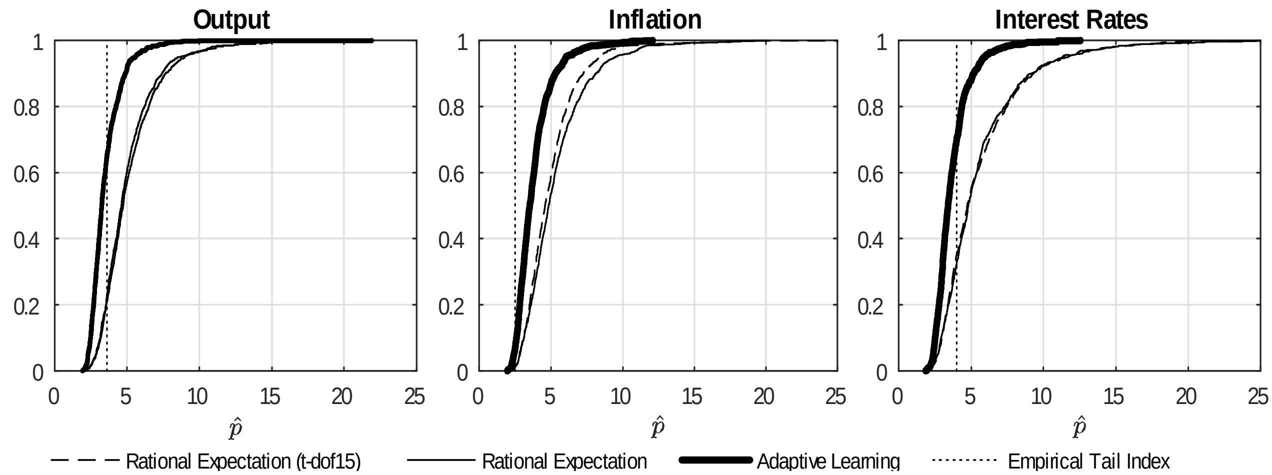


Figure 6. Simulated CDFs.

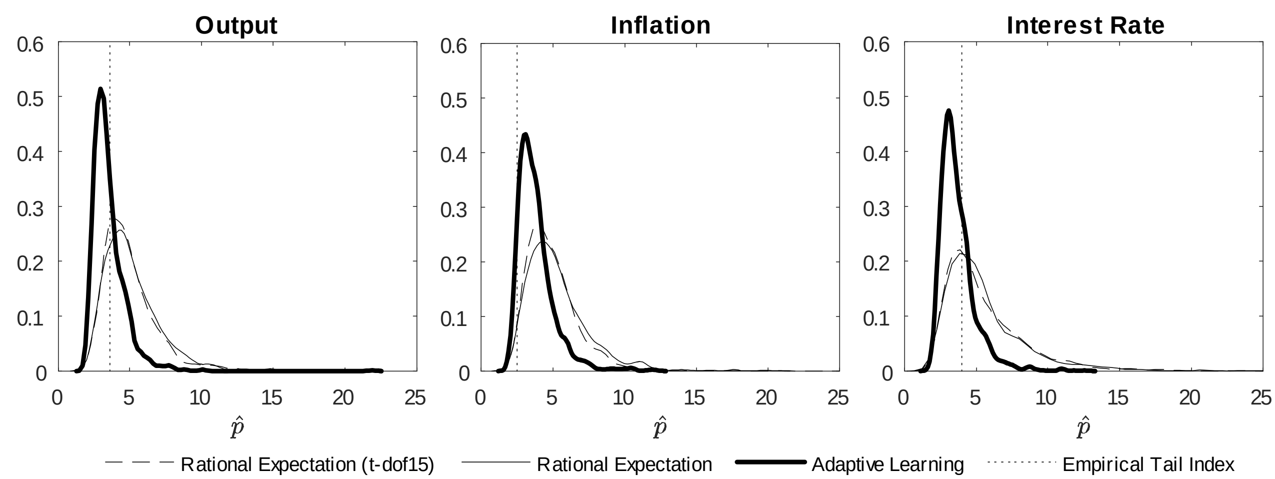


Figure 7. Simulated PDFs.

Figure 6 (Figure 8) and Figure 7 (Figure 9), respectively provide the cumulative and probability distribution function of the tail indices (p's) estimated for each of the 1000 simulated series under the three alternatives described above. For the cumulative distribution function (CDF), the figure captures the probability that the p takes a value less than or equal to p. The dotted line shows the estimated p from the actual data. Two observations are clear from Figure 6. First, the probability of p taking a value closer to the p estimated from the data, is always higher under the adaptive learning model than under other alternatives. Second, for each output, inflation and interest rate, the estimates of p from other alternatives first order stochastically dominates the estimates of p under adaptive learning. Due to stochastic dominance, it is clear that for each of the 1000 series of output, inflation and interest rates, the estimated p is always smaller (hence simulated data depicts fatter tails) under adaptive learning. Figure 7 illustrates the same results using the probability distribution function and shows that the average p under adaptive learning is closer to the p in the data. Furthermore, in Table 6. (Table 7.) Kolmogorov-Smirnov test statistics at 1 percent significance level rejects the null hypothesis that the two samples (from our model versus other alternatives) are drawn from the same distribution. The illustrations suggest that tail indices of our simulated series under adaptive learning come closer to $\hat{p}$ estimate of the actual data relative to the distribution of the p's in the simulated series under other alternatives.

Table 6. Kolmogorov-Smirnov Test (p-value and Test decision (H))

	AL and RE		AL and RE ($t-$dof15)		RE and RE ($t-$dof15)	
	p-value	H	p-value	H	p-value	H
y_t	0.0000	1	0.0000	1	0.2575	0
π_t	0.0000	1	0.0000	1	0.0091	1
r_t	0.0000	1	0.0000	1	0.4931	0

Table 7. Kolmogorov-Smirnov Test (p-value and Test decision (H))

	AL and RE		AL and RE ($t-$dof3)		RE and RE ($t-$dof3)	
	p-value	H	p-value	H	p-value	H
y_t	0.0000	1	0.0000	1	0.0000	1
π_t	0.0000	1	0.0036	1	0.0000	1
r_t	0.0000	1	0.0000	1	0.0000	1

Figure 8 and 9 show similar results as Figure 6 and 7 for output and the interest rate. For inflation, however, the Rational Expectations (t-dof3) model is able to deliver fat tails in line with the data and closer to the tails delivered by Adaptive Learning model but the channel is as discussed, exogenous. The Adaptive Learning model on the other hand provides an endogenous channel for fat tails which originates via a behavioral intuition based on the process of expectations formation by an agent with limited cognitive ability. The agent reduces the sample of data to form expectations which results in deviations from the true values of the variables of interest. Such deviations accumulate over time and manifest themselves as a large deviations and fatter tails.

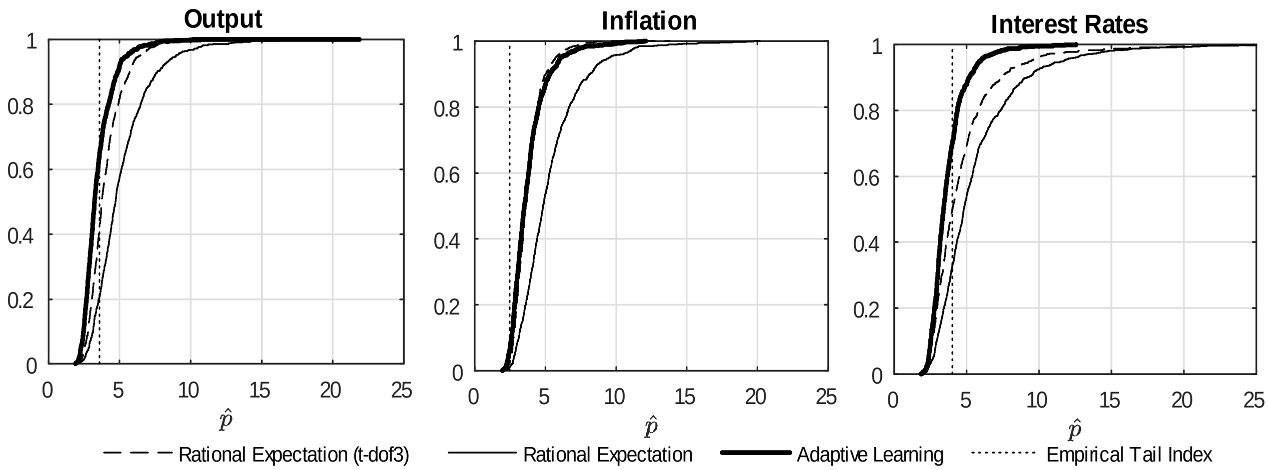


Figure 8. Simulated CDFs.

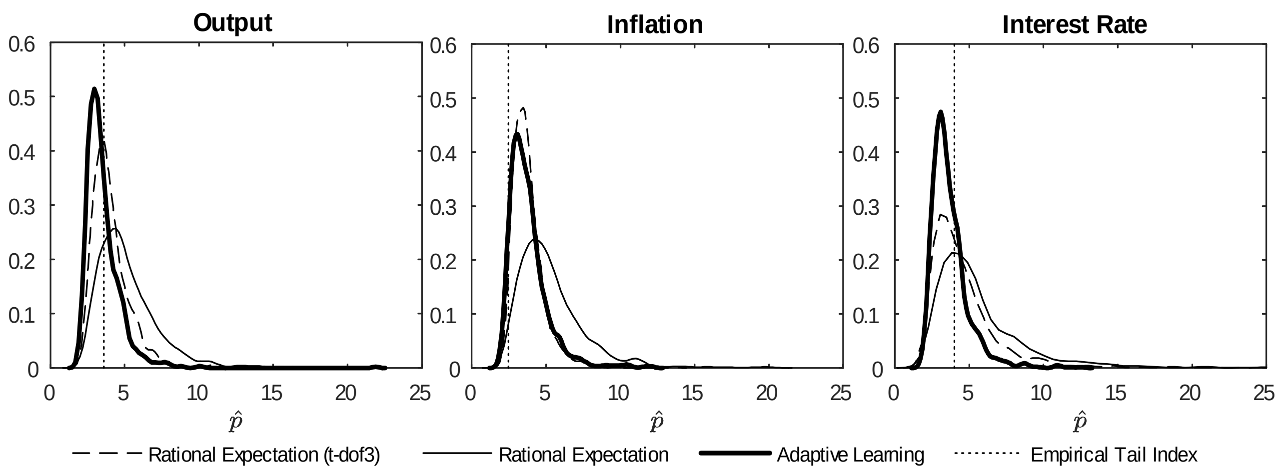


Figure 9. Simulated PDFs

7. Conclusion

In accounting for large deviations of macroeconomic series from their respective trends two approaches are available. The first assumes non-Normal distributions for innovations to structural shocks and has seen some success in being validated by the data. The second proposes an endogenous mechanism, adaptive learning, which delivers a model representation that entails multiplicative and additive noise. We empirically validate this mechanism and find that it can match observed fat tails provided that the gain parameter which governs the extent of learning is larger than usually assumed in the literature. Moreover, a higher gain is needed to reconcile learning algorithms with underlying expectations is a finding validated in survey data by Malmendier and Nagel (2013). This result suggests that the representation of macroeconomic data and models as fixed coefficient VARs may not be as useful as representations in terms of variable coefficient VARs. This finding is similar in spirit to the debate implicit in Sims (2001) and Cogley and Sargent (2001, 2005) who were concerned more with the representation that best suits inflation; our result however is driven by recognizing that constant gain learning naturally leads to an empirical specification that features multiplicative noise.

Various future questions can be investigated. First, can one disentangle the existence of fat tails with the existence of time variation in the system? One example relates to inflation (the series for which it seems that the evidence of fat tails is larger). Is the behavior of inflation better described by fat tails, or is it the outcome of a model in which monetary policy varies from passive to active policy in the middle of the sample. Second, can the large gain as documented in the paper potentially lead the model to fail in other dimensions (maybe in overestimating the volatility of series or underestimating their autocorrelation)? This is beyond the scope of the paper, but an interesting future avenue. Third, econometrically one can complement our results with a fixed point analysis between the linear recursion with multiplicative noise and the time varying state space representation, that determines how agents estimate the constant gain. Given the multivariate nature of our framework, to find the long-run value of the gain that determines the fat tail indices under adaptive learning is a potential future topic. Our simulation exercise indicates a monotonically decreasing relationship between the gain and the possibility of fat tails in macroeconomic aggregates. We conclude therefore that adaptive learning representations can lead to regularly occurring large deviations in macroeconomic systems, and match reality, as our data and analyses suggest.

References

[1] Adam, K. and Woodford, M., 2012. Robustly optimal monetary policy in a microfounded New Keynesian model. Journal of Monetary Economics, 59(5), pp.468-487.

[2] Arnold, B. C., 1983, Pareto Distributions (International Cooperative Publishing House, Fairland, Maryland).

[3] Ascari. G., Fagiolo, G., and A. Roventini, 2015. Fat-Tail Distributions and Business-Cycle Models, Macroeconomic Dynamics, v. 19 (2), pp. 465-476.

[4] Auerbach, A.J. and Gorodnichenko, Y., 2012. Fiscal multipliers in recession and expansion. In Fiscal Policy after the Financial crisis (pp. 63-98). University of Chicago press.

[5] Baxter, M. and R. G. King, 1999. Measuring Business Cycles: Approximate Band-Pass Filters For Economic Time Series, The Review of Economics and Statistics, v. 81 (4), pp. 575-593.

[6] Ben-David, D., Lumsdaine, R.L. and Papell, D.H., 2003. Unit roots, postwar slowdowns and long-run growth: evidence from two structural breaks. Empirical Economics, 28(2), pp.303-319.

[7] Benhabib, J. and C. Dave, 2014, Learning, Large Deviations and Rare Events, Review of Economic Dynamics, v. 17, pp. 367-382.

[8] Bernanke, B. and I. Mihov, 1998a. The liquidity effect and long-run neutrality, Carnegie-Rochester Conference Series on Public Policy v. 49, pp. 149–194.

[9] Bernanke, B. and I. Mihov, 1998b. Measuring monetary policy, Quarterly Journal of Economics, v. 113, pp. 869-902.

[10] Blanchard, O. and Simon, J., 2001. The long and large decline in US output volatility. Brookings papers on economic activity, 2001(1), pp.135-164.

[11] Blume, A., J. Duffy and T. Temzelides, 2010. Self-organized criticality in a dynamic game, Journal of Economic Dynamics and Control, v. 34 (8), pp. 1380-1391.

[12] Branch, W.A. and Evans, G.W., 2007. Model uncertainty and endogenous volatility. Review of Economic Dynamics, 10(2), pp.207-237.

[13] Canzoneri, M., Collard, F., Dellas, H. and Diba, B., 2016. Fiscal multipliers in recessions. The Economic Journal, 126(590), pp.75-108.

[14] Christiano, L. J. and T. J. Fitzgerald, 2003, The Band Pass Filter. International Economic Review, v. 44, pp. 435-465.

[15] Clauset, A., Shalizi, C. R. and M. E. J. Newman, 2009. Power-law Distributions in Empirical Data, SIAM Review, v. 51 (4), pp. 661-703.

[16] Cúrdia, V., Del Negro, M. and D. L. Greenwald, 2014. Rare Shocks, Great Recessions, Journal of Applied Econometrics, v. 29 (7), pp. 1031-1052.

[17] Chib S. and S. Ramamurthy, 2014. DSGE Models with Student-t errors, Econometric Reviews, v. 33, pp. 152-171.

[18] Christiano, L.J., 2007. Comment on On the Fit of NewKeynesian Models by Del Negro, Schorfheide, Smets and Wouters. Journal of Business and Economic Statistics v. 252, pp. 143-151.

[19] Cogley, T. and T. J. Sargent, 2001. Evolving Post World War II U.S. Inflation Dynamics, NBER Macroeconomics Annual,16.

[20] Cogley, T. and T. J. Sargent, 2005. Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII US., Review of Economic Dynamics, v. 8 (2), pp. 262-302.

[21] Deak, S., Levine, P. and Yang, B., 2015, July. A New Keynesian Behavioural Model with Individual Rationality and Heterogeneous Agents. In CEF2015 Conference, June.

[22] De Grauwe, P., 2012. Booms and busts in economic activity: A behavioral explanation. Journal of Economic Behavior & Organization, 83(3), pp.484-501.

[23] DeJong, D. N. and C. Dave, 2011. Structural Macroeconometrics, 2^{nd} Ed., *Princeton University Press.*

[24] Duffy, J., 2012 Macroeconomics: A Survey of Laboratory Research. University of Pittsburgh.

[25] Evans, G.W. and S. Honkapohja, 2001. Learning and Expectations in Macroeconomics. Princeton University Press.

[26] Fagiolo, G., Napoletano, M., Piazza, M. and A. Roventini, 2009. Detrending and the distributional properties of U.S. output time series, Economics Bulletin, v. 29, pp. 3155-3161.

[27] Ferraresi, T., Roventini, A. and Fagiolo, G., 2015. Fiscal policies and credit regimes: a TVAR approach. Journal of Applied Econometrics, 30(7), pp.1047-1072.

[28] Franta, M., 2015. Rare Shocks vs. Non-linearities: What Drives Extreme Events in the Economy? Some Empirical Evidence (No. 2015/04).

[29] Gaspar, V., Smets, F. and Vestin, D., 2006. Adaptive learning, persistence, and optimal monetary policy. Journal of the European Economic Association, 4(2-3), pp.376-385.

[30] Heemeijer, P., Hommes, C., Sonnemans, J. and Tuinstra, J., 2012. An experimental study on expectations and learning in overlapping generations models. Studies in Nonlinear Dynamics & Econometrics, 16(4).

[31] Jaimovich, N. and Rebelo, S., 2007. Behavioral theories of the business cycle. Journal of the European Economic Association, 5(2-3), pp.361-368.

[32] King, Robert G., 2000: The New IS-LM Model: Language, Logic, and Limits, Federal Reserve Bank of Richmond Economic Quarterly, 86(3), 45-103.

[33] Kim, C.J., Nelson, C.R. and Piger, J., 2004. The less-volatile US economy: a Bayesian investigation of timing, breadth, and potential explanations. Journal of Business & Economic Statistics, 22(1), pp.80-93.

[34] Kesten, H., 1973. Random difference equations and renewal theory for products of random matrices, Acta Mathematica, v. 131 (1), pp. 207-248.

[35] Lütkepohl, H., 2007. New introduction to multivariate time series analysis, Berlin and Heidelberg: Springer.

[36] Lubik, Thomas A., and Frank Schorfheide. 2004. Testing for Indeterminacy: An Application to U.S. Monetary Policy. American Economic Review, 94(1): 190-217.

[37] Marcet, A. and J. P. Nicolini, 2005. Money and Prices in Models of Bounded Rationality in High-Inflation Economies, Review of Economic Dynamics, Elsevier, v. 8, pp. 452-479.

[38] Marcet, A. and T. J. Sargent, 1989. Convergence of least squares learning mechanisms in self-referential linear stochastic models, Journal of Economic Theory, Elsevier, v. 48 (2), pp. 337-368.

[39] Malmendier, U. and S. Nagel, 2013. Learning from Inflation Experiences, Unpublished Manuscript.

[40] Massaro, D., 2013. Heterogeneous expectations in monetary DSGE models. Journal of Economic Dynamics and Control, 37(3), pp.680-692.

[41] McConnell, M.M. and Perez-Quiros, G., 2000. Output Fluctuations in the United States: What Has Changed since the Early 1980's?. American Economic Review, 90(5), pp.1464-1476.

[42] Müller, U. K., 2013. Risk of Bayesian inference in misspecified models, and the sandwich covariance matrix. Econometrica, v. 81 (5), pp. 1805-1849.

[43] Milani, F., 2011. Expectation Shocks and Learning as Drivers of the Business Cycle, The Economic Journal, v. 121, pp. 379-401.

[44] Milani, F., 2005. Adaptive Learning and Inflation Persistence, University of California, Irvine-Department of Economics.

[45] Milani, F., 2007. Expectations, learning and macroeconomic persistence. Journal of monetary Economics, 54(7), pp.2065-2082.

[46] Milani, F., 2014. Learning and time-varying macroeconomic volatility. Journal of Economic Dynamics and Control, 47, pp.94-114.

[47] Monache, D and I. Petrella, 2014. Adaptive Models and Heavy Tails, No. 1409. Birkbeck, Department of Economics, Mathematics & Statistics.

[48] Orphanides, A. and Williams, J.C., 2005. Inflation scares and forecast-based monetary policy. Review of Economic Dynamics, 8(2), pp.498-527.

[49] Sargent, T.J., 1993. Bounded Rationality in MacroeconomicsOxford University Press.

[50] Sargent, T. J., 1999. The Conquest of American Inflation, Princeton, NJ: Princeton.

[51] Sims, C. A., 2002. Solving Linear Rational Expectations Models, Computational Economics, v. 20 (1), pp. 1-20.

[52] Smets, F. and R. Wouters, 2007. Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach, The American Economic Review, v. 97 (3), pp. 586-606.

[53] Sims, C.A., 1980. Comparison of interwar and postwar business cycles: monetarism reconsidered. American Economic Review, 250–257.

[54] Sims, C.A., 1999. Drifts and breaks in monetary policy. Unpublished manuscript. Department of Economics, Princeton University.

[55] Sims, C.A., 2001. Comment on Sargent and Cogley's Evolving Post-WorldWar II US Inflation Dynamics. NBER Macroeconomics Annual 16, 373–379.

[56] Stock, J. H., 2001. Discussion of Cogley and Sargent,'Evolving Post-World War II Inflation Dynamics', NBER Macroeconomics Annual, v. 16.

[57] Stock, J. H., and M. W. Watson, 2012, Disentangling the Channels of the 2007-2009 Recession, Brookings Papers on Economic Activity Spring 2012, pp. 81-135.

[58] Stock, J.H. and Watson, M.W., 1999. Forecasting inflation. Journal of Monetary Economics, 44(2), pp.293-335.

[59] Stock, J.H. and Watson, M.W., 2003. Has the business cycle changed? Evidence and explanations. Monetary policy and uncertainty: adapting to a changing economy, pp.9-56.

[60] Schneider, W., 2007. Analytical uses of Kalman filtering in econometrics-A survey, Statistical Papers v. 29 (1), pp. 3-33.

[61] Woodford, Michael. Interest and prices: Foundations of a theory of monetary policy. Princeton, NJ: Princeton University Press, 2003.

Appendix: Estimation Procedure

In this Appendix we detail the method used in our minimum distance estimation of a vector of deep parameters (μ). First, denote a vector of empirical targets as $\varkappa$. In our case this vector can consist solely of tail indices from Table 1. (or tail indices in addition to implied moments of the tail of the stationary distribution of the series in question presented in the Appendix: Additional Analyses). For example, suppose we wish to only match, by choice of μ, the tail index of HP filtered output (≈ 4 from Table 1.). Then the empirical target ($\varkappa$) is just 4. If we wish to match the tail index of HP filtered output and investment (whose tail index is ≈ 3 from Table 1.) then the empirical target is the vector $\varkappa = [4, 3]'$.

1. Having set an empirical target vector $\varkappa$ we next use a Kalman smoother to obtain its model counterpart, as follows. Given a parametrization for μ, the model, written in terms of deviations from trend so that HP filtered data can be used, in state space form, is

$$b_t = b_{t-1} + gX_{t-1}(X_t - X'_{t-1}b_{t-1}), \ b_0 \text{ given}, \tag{51}$$

$$X_t = \Gamma b^2_{t-1}X_{t-1} + \Psi U_t, \tag{52}$$

$$Y_t = H'X_t. \tag{53}$$

   where X_t denotes model variables, Y_t denotes HP filtered data and the deep parameters μ consist of the constant gain g along with the elements that appear in the matrices Γ and Ψ; H' is simply the usual selection matrix for the observer equation that links observed HP filtered data Y_t to its model counterpart X_t. Note that the state equation is a time varying one, so that this is a time varying state space representation. Thus, given a parameterization for μ the above delivers a time varying state space representation of the model. Next, we obtain the smoothed values of the state vector ($\widehat{X}_t$) via a Kalman smoother. We do not claim that this is optimal, it is simply a tool to obtain $\widehat{X}_t$ (see the references to Lütkepohl (2007) and Schneider (2007) in Section 4.1).

2. Once the smoothed values $\widehat{X}_t$ are in hand for a given parameterization of μ we have a set of series for which we can calculate model counterparts of the empirical targets ($\varkappa$) which we denote as $\varkappa(\mu)$. When these empirical targets are tail indices we use the method of Clauset et al. (2009) to calculate the tail index of the smoothed series ($\widehat{X}_t$) corresponding to a parametrization. There is one twist in this procedure for our model which is built to explain the tail behavior of output, inflation and interest rates. Suppose that the empirical target is the tail index for HP filtered output. Then in the observer equation we only use data on the remaining two series (inflation and interest rates) otherwise we would be matching the tail index of output exactly. In order to let the procedure tell us what is the appropriate tail index for output, we cannot include output in the observer equation.

3. Given a parametrization of μ we have in hand the column vector of empirical targets $\varkappa$ and the column vector of their model counterpart $\varkappa(\mu)$, we next search over the parameter space to minimize the squared difference between $\varkappa$ and $\varkappa(\mu)$ in order to

estimate values for μ; that is, our estimates are delivered by

$$\min_{\mu} F = [\varkappa - \varkappa(\mu)]'[\varkappa - \varkappa(\mu)] \tag{54}$$

4. Standard errors computed using the Hessian of the above objective function at the parameter estimates (see DeJong and Dave,2011). It is critical to note that this exercise is *not* a moment matching exercise in the conventional sense. The vector $\varkappa$ can contain tail indices as well as implied moments of the *tail of the stationary distribution* of a set of series. We make no claims as to the matching of moments, simulated or otherwise, of the distribution of the data given our exclusive focus on the LRMN assumption on the DGP and the associated model characterization of the tail of the stationary distribution of macroeconomic series.

Note that if in step 2 we instead wish to match the tail indices of HP filtered output and the associated moments of the tail of the stationary distribution of HP filtered output (see Appendix: Additional Analyses), then the vector $\varkappa$ would consist of the tail index of HP filtered output (4) followed by the values of the first three moments of HP filtered output. The rest of the steps then follow as mentioned above.

Appendix: Additional Analyses

The tail index summarizes the heaviness of the tail, and characterizes also the existence of finite moments of the entire distribution. In particular, the relation between the tail index ($\widehat{p}$) and the existence of the number of moments where n denotes that number is:

$$n = \widehat{p} - 1. \tag{55}$$

Higher moments such that $n > \widehat{p} - 1$ are infinite and therefore do not exist. Recall from Table 1 that the tail indices for HP-filtered output, inflation and interest rates are as follows: $p^y = 3.3663$, $p^\pi = 2.5105$ and $p^r = 4.1979$. As a result the number of moments that are finite for output, inflation and interest rate, respectively is $n^y = 2$, $n^\pi = 1$ and $n^r = 3$. Letting m denote the exact value of a moment, this implies that only first two moments i.e., the mean (m_1^y) and standard deviation (m_2^y), exist for output, only the first moment i.e., the mean (m_1^π) exist for inflation and the first three moments i..e., mean (m_1^r), standard deviation (m_2^r) and kurtosis (m_3^r) exist for the interest rate. Using this information we re-estimate our parameters by using an objective function that consists only of the number of finite moments given the values of the tail indices in the data. Corresponding objective functions for this exercise therefore are as follows:

$$\min_{\mu} F = \left[\begin{pmatrix} m_1^y \\ m_2^y \end{pmatrix} - \begin{pmatrix} \widehat{m}_1^y(\mu) \\ \widehat{m}_2^y(\mu) \end{pmatrix} \right]' \left[\begin{pmatrix} m_1^y \\ m_2^y \end{pmatrix} - \begin{pmatrix} \widehat{m}_1^y(\mu) \\ \widehat{m}_2^y(\mu) \end{pmatrix} \right], \tag{56}$$

$$\min_{\mu} F = \left[m_1^\pi - \widehat{m}_1^\pi(\mu) \right] \left[m_1^\pi - \widehat{m}_1^\pi(\mu) \right]', \tag{57}$$

$$\min_{\mu} F = \left[\begin{pmatrix} m_1^r \\ m_2^r \\ m_3^r \end{pmatrix} - \begin{pmatrix} \widehat{m}_1^r(\mu) \\ \widehat{m}_2^r(\mu) \\ \widehat{m}_3^r(\mu) \end{pmatrix} \right]' \left[\begin{pmatrix} m_1^r \\ m_2^r \\ m_3^r \end{pmatrix} - \begin{pmatrix} \widehat{m}_1^r(\mu) \\ \widehat{m}_2^r(\mu) \\ \widehat{m}_3^r(\mu) \end{pmatrix} \right]. \tag{58}$$

Parameter estimates are provided in Table 8. Our main result is that the estimated g is smaller than the estimated g provided in Table 3 and 4 where we matched the tail indices alone. However, the estimate is still considerably larger than the existing literature. Other parameters are also somewhat different compared to Milani (2011). This highlights that the higher estimate of g is not just a function of the objective function we aim to minimize since generally speaking, g is always larger than the g estimate documented in existing research. We provide the sensitivity of our objective function to various values of g in Figure 10. These figures below provide illustrative evidence that larger values of g are the ones successful in matching model with data.[10]

[10] Furthermore, the function values in the illustration are deceptive to the naked eye as the function seems to reach a minimum of zero but such instances are very close to zero and not actually zero.

Table 8. Matching Moments

	Milani	y_t Est.	π_t Est.	r_t Est.
γ_1	1.4170	1.2244	1.3770	1.5801
		(0.0001)	(0.0001)	(0.0001)
γ_2	0.2210	0.4987	0.2158	0.4211
		(0.0001)	(0.0001)	(0.0001)
θ	0.9498	0.9758	0.9619	0.8806
		(0.0011)	(0.0001)	(0.0001)
κ	0.0350	0.0180	0.0370	0.0419
		(0.0007)	(0.0001)	(0.0001)
τ	0.2360	0.5719	0.2478	0.3707
		(0.0003)	(0.0001)	(0.0001)
ρ_1	0.3538	0.0127	0.2888	0.2746
		(0.0001)	(0.0001)	(0.0001)
ρ_2	0.1746	0.3199	0.1661	0.4728
		(0.0001)	(0.0001)	(0.0001)
	N/A	0.0619	0.1984	0.3860
		(0.0001)	(0.0001)	(0.0001)
β	0.9615	0.9304	0.9660	0.7945
		(0.0002)	(0.0001)	(0.0001)
g	0.0196	0.0559	0.0447	0.0327
		(0.0009)	(0.0001)	(0.0001)
σ_1	0.7700	0.4749	0.6230	0.6246
		(0.0009)	(0.0001)	(0.0001)
σ_2	0.2970	0.3751	0.2618	0.0633
		(0.0007)	(0.0001)	(0.0001)
σ_3	0.2070	0.2395	0.2267	0.0843
		(0.0007)	(0.0001)	(0.0001)
F		1.2865E-09	5.0936E-11	0.1098

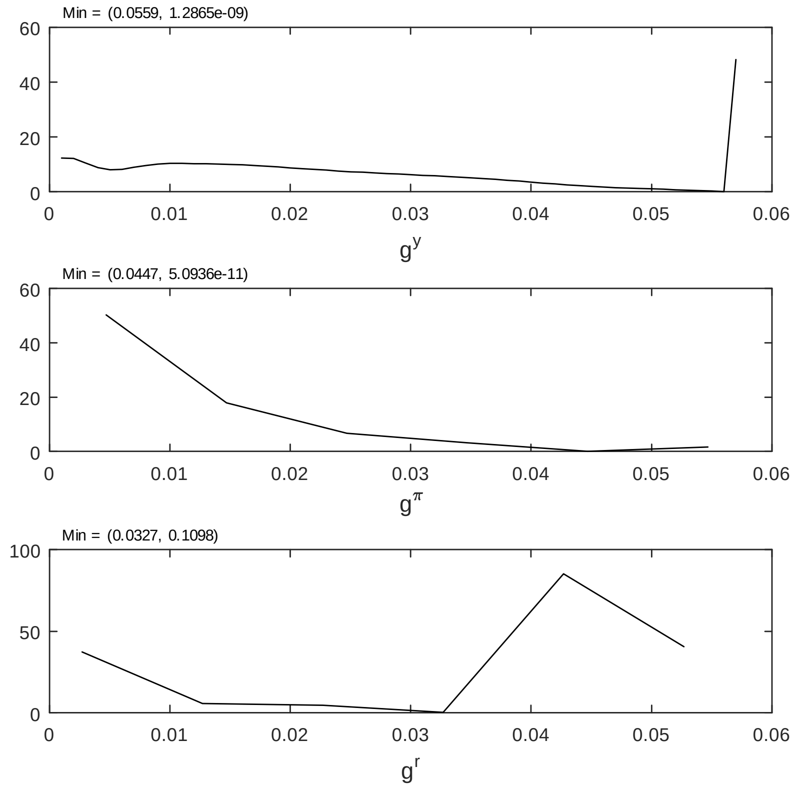


Figure 10.

We next combine the objective function specifications from the last two sections. In particular, we use both the tail index and the associated moments of the data, as follows:

$$\min_{\mu} F = \left[\begin{pmatrix} p^y \\ m_1^y \\ m_2^y \end{pmatrix} - \begin{pmatrix} \widehat{p}^y(\mu) \\ \widehat{m}_1^y(\mu) \\ \widehat{m}_2^y(\mu) \end{pmatrix} \right]' \left[\begin{pmatrix} p^y \\ m_1^y \\ m_2^y \end{pmatrix} - \begin{pmatrix} \widehat{p}^y(\mu) \\ \widehat{m}_1^y(\mu) \\ \widehat{m}_2^y(\mu) \end{pmatrix} \right], \tag{59}$$

$$\min_{\mu} F = \left[\begin{pmatrix} p^\pi \\ m_1^\pi \end{pmatrix} - \begin{pmatrix} \widehat{p}^\pi(\mu) \\ \widehat{m}_1^\pi(\mu) \end{pmatrix} \right]' \left[\begin{pmatrix} p^\pi \\ m_1^\pi \end{pmatrix} - \begin{pmatrix} \widehat{p}^\pi(\mu) \\ \widehat{m}_1^\pi(\mu) \end{pmatrix} \right], \tag{60}$$

$$\min_{\mu} F = \left[\begin{pmatrix} p^r \\ m_1^r \\ m_2^r \\ m_3^r \end{pmatrix} - \begin{pmatrix} \widehat{p}^r(\mu) \\ \widehat{m}_1^r(\mu) \\ \widehat{m}_2^r(\mu) \\ \widehat{m}_3^r(\mu) \end{pmatrix} \right]' \left[\begin{pmatrix} p^r \\ m_1^r \\ m_2^r \\ m_3^r \end{pmatrix} - \begin{pmatrix} \widehat{p}^r(\mu) \\ \widehat{m}_1^r(\mu) \\ \widehat{m}_2^r(\mu) \\ \widehat{m}_3^r(\mu) \end{pmatrix} \right]. \tag{61}$$

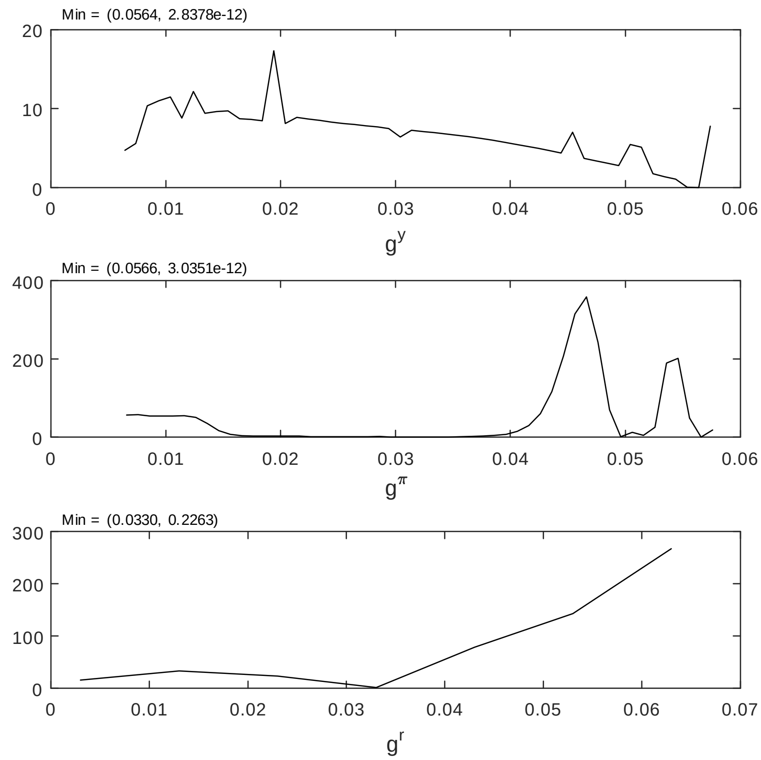
We present our results in Table 9. We again find similar results for estimated parameter g when compared to the estimates from the last two tables. Specifically, the objective function that minimizes jointly the tail index and the moments for our variable of interest results in considerably different parameter estimates that govern transmission. This means that under this objective function, a higher g and a different transmission mechanism delivers the fat tails. We provide the sensitivity of our objective function to various values of g in Figure

11. These figures below provide illustrative evidence that the larger values of g are the ones successful in matching model with data.[11]

Table 9. Matching Tail Coefficients and Moments

		y_t	π_t	r_t
	Milani	Est.	Est.	Est.
γ_1	1.4170	1.5775	1.4790	1.4454
		(0.0001)	(0.0010)	(0.0001)
γ_2	0.2210	0.3252	0.1605	0.2564
		(0.0001)	(0.0210)	(0.0001)
θ	0.9498	0.9375	0.9663	0.9115
		(0.0014)	(0.0153)	(0.0001)
κ	0.0350	0.0004	0.0326	0.0387
		(0.0028)	(0.1590)	(0.0001)
τ	0.2360	0.3729	0.1964	0.2708
		(0.3515)	(0.0748)	(0.0001)
ρ_1	0.3538	0.3526	0.3767	0.3645
		(0.0003)	(0.0826)	(0.0001)
ρ_2	0.1746	0.3585	0.2046	0.1500
		(0.0001)	(0.0126)	(0.0001)
ρ_3	N/A	0.0078	0.2657	0.2187
		(0.0001)	(0.0016)	(0.0001)
β	0.9615	0.9582	0.9346	0.8554
		(0.0001)	(0.1018)	(0.0001)
g	0.0196	0.0564	0.0566	0.0330
		(0.0029)	(0.0405)	(0.0001)
σ_1	0.7700	0.6496	0.7976	0.7406
		(0.0010)	(0.0002)	(0.0001)
σ_2	0.2970	0.2683	0.2453	0.3361
		(0.0035)	(1.5253)	(0.0001)
σ_3	0.2070	0.1644	0.3754	0.2263
		(0.0011)	(0.0920)	(0.0001)
F		2.8378E-12	3.0351E-11	0.2024

[11]Furthermore, the function values in the illustration are deceptive to the naked eye as the function seems to reach a minimum of zero but such instances are very close to zero but not actually zero.



Min = (0.0564, 2.8378e-12)
20
10
0
0
0.01
0.02
0.03
0.04
0.05
0.06
g^y
Min = (0.0566, 3.0351e-12)
400
200
0
0
0.01
0.02
0.03
0.04
0.05
0.06
g^π
Min = (0.0330, 0.2263)
300
200
100
0
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
g^r

Figure 11.