# A Tale of Fat Tails*

Chetan Dave† NYUAD

Samreen Malik‡ NYUAD

September 4, 2017

**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.

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

*We especially thank editor Eric M. Leeper, associate editor & anonymous referees for many comments that helped us significantly improve the paper. We extend special thanks to Jess Benhabib for his guidance, comments and suggestions throughout the process. We also thank Christian Haefke, Jean Imbs, John Leahy, Max Mihm, Romain Ranciere, and seminar participants at the Midwest Macroeconomics meetings, 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.[1] We show that a DSGE model under adaptive learning (AL) 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 fluctuations. 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) AL that delivers fat tails for endogenous variables in an otherwise standard model.

AL 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 AL 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 AL agents revise their forecast rules in response to incoming data so as to ascertain that stochastic process over time.[2] 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

[1] We use the term large fluctuations to mean that the frequency of realizations that deviate substantially from the trend are higher than the frequencies of the same realizations under a Normal distribution.

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

multiplicative and additive noise (LRMN) written as

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

and with a stationary distribution for $X_t$ different than that of a VAR due to the interplay between the *stochastic* multiplicative term ($\Phi_t$) and the stochastic additive term ($\varepsilon_t$).[3] 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. 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.[4]

Since under SGCG learning a 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 conduct a formal test to establish whether the hypothesized power-law is a plausible fit to the data and our test statistics show that data that enter a DSGE empirical exercise are usually not Normal.

We find that data exhibit characteristics consistent with a LRMN assumption on the DGP. In particular, the 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.[5] A finding of a higher constant gain does not violate any theoretical or empirical requirement that $g$ be near zero (so an AL model is in a small vicinity of its RE solution). 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 AL in terms of being able to replicate 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 AL is enough to account

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

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

[5] Malmendier and Nagel (2013) find evidence in favor of a constant gain (approximately 0.02) using survey data on expectations across generations.

for observed fat tails without altering the model's fit to other distributional dimensions of the data.

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 model with a 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 fluctuations, 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

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

where $\theta$ denotes a vector of parameters, $X_t$ denotes all model variables (in logarithmic deviations from respective steady state values) and $\varepsilon_t$ 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 ($\eta_t$), one then obtains the form employed by Sims (2001)

$$\Gamma_0(\theta)X_t = \Gamma_1(\theta)X_{t-1} + \Psi(\theta)\varepsilon_t + \Pi(\theta)\eta_t. \tag{4}$$

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

$$X_t = F(\theta)X_{t-1} + G(\theta)\varepsilon_t, \tag{5}$$

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, yields the state-space empirical representation of a linear DSGE model under RE. Specifying, say, a Normal distribution for the innovations alongside a Kalman Filter allows one to write down the likelihood function ($L(Y_t|\theta)$) employed in estimating parameters using, for example, either Maximum Likelihood or Bayesian full-information techniques. Given parameter estimates, the state equation is used to compute the usual objects of inquiry (e.g. impulse responses, variance decompositions etc.).[6] The key to such an empirical exercise is data-model congruency: the assumption on the underlying DGP is that macroeconomic

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

data in their cyclic form follow a fixed-coefficient VAR

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

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 $\upsilon_t$ 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 in $Y_t$ (e.g., rare but deep recessions or booms) one can assume a fat-tailed distribution for $\upsilon_t$ (and correspondingly for $\varepsilon_t$) 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 $\upsilon_t$ (and correspondingly for $\varepsilon_t$), or 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 $u_t$ *and* $\Phi_t$ are both 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 to endogenously generate large movements in $Y_t$ in response to thin tailed shocks one assumes 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 SGCG AL, leads to a model representation being of the LRMN form.

Following a large literature described in Evans and Honkapohja (2001), under AL agents are presumed to only employ information up until time $t-1$ in forming the required expectations.[7] This alters the original specification of the model to

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

[7] The timing assumption contains the information set on which an agent conditions her expectations. For alternative timing assumptions (e.g. contamporenous timing), see Marcet and Sargent (1989).

Under AL agents estimate their model of the dynamics of economic variables, called the perceived law of motion (PLM), by recursive least squares, arguably the most common estimation method in econometrics. The PLM has the same functional form as the RE equilibrium but possibly different coefficients ($b_t$) since agents do not know the RE equilibrium. To estimate the PLM, agents use past data and then generate forecasts using the estimated model. Thus, a 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[8]

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

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 fluctuations in $b_t$ would then drive the same in $X_t$. However this raises analytical issues since the driving processes ($U_t$) and innovations ($\varepsilon_t$) 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 $\varepsilon_t$ having inserted the ALM into the SGCG learning rule.[9] 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 since the RE solution takes a VAR form. Instead let's assume that the PLM is a time varying moving average (MA) process

$$X_t = \hat{\Phi}_t U_t. \tag{13}$$

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

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

[8] Evans et. al (2010) extend the analysis of econometric learning to settings where agents allow for parameter drift or parameter uncertainty in beliefs. With these features, they demonstrate that the SGCG rule is a convenient and a natural way to model learning. Intuitively, if a true DGP of the economic environment is a random walk, then it makes sense to weight recent observations more heavily (i.e. using a constant gain rule for learning) than a decreasing gain formulation. This is because in such an environment recent observations have more relevant information for forecasting the next realization.

[9] For univariate cases see Benhabib and Dave (2014).

The evolution of $\hat{\Phi}_t$ is governed by the learning rule

$$\hat{\Phi}_t = \hat{\Phi}_{t-1} + gU_{t-1}(X_t - \hat{\Phi}_{t-1}U_{t-1}) \tag{15}$$

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

$$\hat{\Phi}_t = \left[I + gA(\theta)U_{t-1}U_t - gU_{t-1}U_{t-1}\right]\hat{\Phi}_{t-1} + gB(\theta)U_{t-1}U_t. \tag{16}$$

The SGCG AL model can therefore be written as

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

Next we recognize that the VAR process for structural errors can always be written in its Wold representation ($P(L)\varepsilon_t$) yielding the LRMN system

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

In this representation the multiplicative term $\Lambda_t$ in equation (22) is now only a function of the innovations $\varepsilon_t$ in equation (23) suggesting that (following Kesten, 1973) the tail of the stationary distribution of the coefficients $\hat{\Phi}_t$ may follow a power law. This would then impart fat tails for $X_t$ as well since $X_t = \hat{\Phi}_t U_t$.[10] To show that the tail of the stationary distribution of model variables $X_t$ is fat as the constant gain 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 AL) and given that SGCG AL 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 it may be the case that history repeats itself due to short memories and macroeconomic variables exhibit "predictably rare" but large fluctuations from trend.

[10] 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).

# 3. Related Literature

Four strands of the existing literature are related to our analyses. First, recent empirical findings suggest that data on macroeconomic variables of interest such as output growth and inflation are seldom Normally distributed. Christiano (2007) provides pre-Great Recession evidence and Ascari et. al. (2015) provides evidence for 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 extreme output growth patterns tend 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 preferred by the data. However, the fat tails delivered under these models could be the result of non-Gaussian distributions assumed for exogenous errors. 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 a result, in the third strand of the literature, Cogley and Sargent (2001, 2005) suggest that attributing adaptive behavior to agents can generate non-linearities observed in data that can manifest themselves as drifting coefficients. Hence they specify time varying VARs in addition to assuming stochastic volatility in innovations. Monache and Paterella (2014) build on the framework of Cogley and Sargent (2001) by replacing Normally distributed shocks with Student's-$t$ shocks.[11] What these analyses 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 New Keynesian (NK) model under learning. Milani (2005) studies whether learning can provide a reasonable source of the observed persistence in inflation and finds that persistence depends on the learning

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

speed. Primiceri (2009) further contributes to the discussion by demonstrating that in models exhibiting 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 considers the role of non-linearity originating from state dependence in DSGE models to understand the dynamics of an economy (e.g., Auerbach and Gorodnichenko, 2012, Ferraresi et. al., 2015, Canzoneri et. al., 2016). Non-linearity is incorporated by the possibility of a regime switch, for e.g., Franta (2015) introduces 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. Massaro (2013) instead exogenously assumes multiple heuristics where agents discretely shift between rules for expectations formation in an AL model.

Similar to the literature 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 (see Kesten (1973), Goldie (1991), Saporta (2005) and Roitershtein (2007)). In this literature the key is the fact that the variable(s) of interest, being modeled as a LRMN, have stationary distributions whose tail is approximated by a power law. The smaller the coefficient on that power law, the fatter the tail. Application of these technical results can also be found in the macroeconomic literature (see e.g., Benhabib (2009), Benhabib et. al., (2011) and Benhabib and Dave (2014)).

# 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. We then describe our data, present the fat tail measures and discuss the interpretation of our documented facts.

## 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 data ($Y$). In particular, data obey 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

and in such cases we say that the tail of the distribution is thick. Estimation of that index proceeds as follows. Taking logarithms of (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 approximated with a histogram of observed $Y$. The above slope determines the thickness of the tail and can be estimated via a least squares linear regression. A steeper negative slope (i.e. a higher $p$) implies that the probability of larger values of $\ln(Y)$ are less frequent compared to a flatter negative slope (i.e. 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. Clauset et. al. (2009) provide a maximum likelihood (ML) procedure that 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 $Y_{\min} \leq Y_i$.[12] The width of the likelihood maximum estimator provides standard errors for $\widehat{p}$, given by[13]

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

Additionally $\widehat{p}-1$ measures the number of moments that exist for the underlying 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 distribution but thinner than that of the power law distribution. Moreover, unlike a power law distribution, the EP is characterized by finite moments of any order.[14]

To investigate if the hypothesized distribution of the data follows a power law, we conduct a *goodness-of-fit-test* from Clauset et. al. (2009). This test is driven by the idea that even if data are drawn from a power law, there will always be some small deviations from that law merely due to the random nature of the sampling process (statistical fluctuations). Therefore, the goodness-of fit-test allows us to distinguish deviations which are due to statistical

[12] 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 ML, and the Kolmogorov-Smirnov goodness-of-fit statistic (denoted by $D$) is calculated; 2) A value of $Y_{min}$, that gives the minimum value $D$ over all values of $Y_{min}$ is then selected as the estimate of $Y_{min}$.

[13] For details of derivation of the likelihood function see Appendix B of Clauset et al. (2009).

[14] Note that one cannot directly compare the estimates of the tail index of the *detrended* data under a power law distribution in our analyses with the EP of *raw* data (e.g. Fagiolo, 2008).

fluctuations versus deviations that arise because the data are drawn from a non-power-law distribution. The cut-off for the $p$-value of this is chosen based on the desired precision level of test statistic. We use a $p$-value larger than 0.1 to indicate that the difference between the empirical data and the model can be attributed to statistical fluctuations only and in contrast a $p$-value closer to zero indicates that the model is not a plausible fit to the data. Therefore estimates of tail indices are meaningful if this test is passed. Appendix A provides further details on the goodness of fit test.

## 4.2. Data

The FRED database provides raw series on output, consumption, investment, prices, population, money stocks and interest rates (GDPC96, PCECC96, GPDIC1, GDPDEF, CNP16OV, M2SL and TB3MS respectively) spanning 1948-2014.[15] With the raw data in hand, per-capita series are constructed as:

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

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

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

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 the set $\{y_t, c_t, i_t, m_t, \pi_t, r_t\}$ can then be detrended using the various methods described in DeJong and Dave (2011) in order to generate cyclic series, denoted as $\{\widehat{y}_t, \widehat{c}_t, \widehat{i}_t, \widehat{m}_t, \widehat{\pi}_t, \widehat{r}_t\}$.

## 4.3. Tail Index Estimation: Results

We first investigate if data detrended with the HP filter are indeed Normal using direct tests (the Anderson-Darling, Shapiro-Wilk, Shapiro-Francia, Jarque-Bera and D'Agostino and Pearson tests) and report the test decision based on the null hypothesis that our sample come from a population with a Normal distribution. A rejection of this null hypothesis is denoted by a 0 whereas a 1 denotes that we are unable to reject the null hypothesis. We conduct these tests at both the 5% and 10% significance levels. Our results, presented in Table 1, show that at a 10% significance level, all tests reject the null hypothesis for output, investment, inflation and interest rate but not for consumption and money. In the last five rows of Table 1, we present the same results at a 5% significance level and we see that test decisions stay consistent for all the variables except output for which we fail to reject the null under the first three tests. Overall, we conclude that there is evidence of non-Normality for output, investment, inflation and interest rates (at a 10% significance level), but not so much for consumption and money.

[15] The last three time series are monthly so we compute monthly averages to obtain data at a quarterly frequency.

Table 1. Normality of Cyclical Components

| Test ($\alpha = 0.1$) | $\widehat{y}_t$ | $\widehat{c}_t$ | $\widehat{i}_t$ | $\widehat{m}_t$ | $\widehat{\pi}_t$ | $\widehat{r}_t$ |
|---|---|---|---|---|---|---|
| Anderson-Darling | 0 | 1 | 0 | 1 | 0 | 0 |
| Shapiro-Wilk | 0 | 1 | 0 | 1 | 0 | 0 |
| Shapiro-Francia | 0 | 1 | 0 | 1 | 0 | 0 |
| Jarque-Bera | 0 | 1 | 0 | 1 | 0 | 0 |
| D'Agostino & Pearson | 0 | 1 | 0 | 1 | 0 | 0 |
| Test ($\alpha = 0.05$) | $\widehat{y}_t$ | $\widehat{c}_t$ | $\widehat{i}_t$ | $\widehat{m}_t$ | $\widehat{\pi}_t$ | $\widehat{r}_t$ |
| Anderson-Darling | 1 | 1 | 0 | 1 | 0 | 0 |
| Shapiro-Wilk | 1 | 1 | 0 | 1 | 0 | 0 |
| Shapiro-Francia | 1 | 1 | 0 | 1 | 0 | 0 |
| Jarque-Bera | 0 | 1 | 0 | 1 | 0 | 0 |
| D'Agostino & Pearson | 0 | 1 | 0 | 1 | 0 | 0 |

Since we have now established that our data for output, inflation and the interest rate are non-Normal, we estimate the tail index for these variables using the procedure described in Section 4.1.[16] For each of these series we report: estimates of its tail index ($\widehat{p}$) along with associated standard errors ($s.e.$) in Table 2. Additionally, the table also provides the $p$-value which is based on a *goodness-of-fit* test as described in Section 4.1. Since Canova (1998) showed that the choice of detrending method can affect business cycle stylized facts, we estimate tail indices under all of the various detrending methods usually employed in DSGE models.

Table 2. Tail Index Estimates For Cyclical Components

| | First Difference | | Linear Trend | | HP-Filter Quarterly | |
|---|---|---|---|---|---|---|
| | $\widehat{p}$ $(s.e.)$ | $p$-value | $\widehat{p}$ $(s.e.)$ | $p$-value | $\widehat{p}$ $(s.e.)$ | $p$-value |
| $\widehat{y}_t$ | 5.1386 (4.1865) | 0.5748 | 3.7068 (8.9334) | 0.0001 | 3.6395 (0.7147) | 0.1963 |
| $\widehat{\pi}_t$ | 2.6608 (0.5990) | 0.289 | 5.0594 (1.941) | 0.5574 | 2.4898 (0.3698) | 0.6086 |
| $\widehat{r}_t$ | 3.1239 (0.9867) | 0.2492 | 3.2615 (0.7651) | 0.0572 | 4.8631 (1.4991) | 0.4289 |
| | CF-Filter | | BK-Filter | | HP-Filter Annual | |
| | $\widehat{p}$ $(s.e.)$ | $p$-value | $\widehat{p}$ $(s.e.)$ | $p$-value | $\widehat{p}$ $(s.e.)$ | $p$-value |
| $\widehat{y}_t$ | 3.3663 (0.4459) | 0.3871 | 4.2914 (2.1432) | 0.2198 | 3.5418 (1.7982) | 0.6890 |
| $\widehat{\pi}_t$ | 2.1743 (0.2365) | 0.0283 | 2.6515 (0.2434) | 0.7541 | 2.2822 ( 1.7989 ) | 0.6157 |
| $\widehat{r}_t$ | 4.116 (2.2901) | 0.0742 | 4.9127 (1.4759) | 0.6415 | 3.4639 (1.7954) | 0.4039 |

Qualitatively, Table 2 shows that for most of the detrending methods, the tail indices of all of our variables of interest are small, pointing towards fatter tails. In particular, these indices (i.e. small $\widehat{p}$'s) provide a strong indication that the assumption of Gaussianity does not hold (consistent with evidence from Table 1), rather a power law is a more plausible fit (i.e. $p$-values greater than 0.1), and there is strong evidence of fat tails in the data.

[16] It is important to note that non-normality as documented in Table 1 does not imply that the distribution of the variable in question has fat tails. The reported test statistics in Table 2 provide evidence of the presence (and significance) of fat tails.

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 standard errors are larger than other detrending methods and therefore the *goodness-of-fit test* does not pass. However, both the estimates and the standard errors of the tail indices for other filtering techniques are quite comparable. Reported $p$-values of our tests show that in almost all cases (barring the CF filter for inflation and the interest rate and a linear trend for output and interest rates), the $p$-values are always bigger than 0.1 indicating that data are approximated well by a power law and therefore the fat tail indices are indeed significant and meaningful. In the empirical implementation and simulation exercises reported below we focus on the tail indices estimated for HP-filtered data. We use the HP-filter $\widehat{p}$ estimates because tests show that the variables for the NK model are indeed power law distributed (for quarterly and yearly data) and that HP filtered data are most commonly employed 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. We therefore also use the implied moments from the tail index as additional empirical targets in our analyses.

# 5. Structural Estimation

## 5.1. Method

We denote a column vector of empirical targets (e.g. tail indices from Table 2) as $\varkappa$ and note that coupling the time varying VAR form of a model's solution with an observer equation (using HP filtered data) allows us to get via a Kalman smoother the smoothed values of the state vector which are used to calculate model targets. Model targets given a parametrization of $\mu$ are denoted by $\varkappa(\mu)$. Targets can be tail indices estimated using Clauset et. al (2009) which we use in Section 5.2 and/or implied moments from tail indices which we use in Section 5.3. For any targets we search over the parameter space to minimize the squared difference between $\varkappa$ and $\varkappa(\mu)$ in order to estimate values for $\mu$; i.e., our estimates are delivered by

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

with standard errors computed using the Hessian of the above objective function at the parameter estimates. 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* innovations 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.[17] 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* $\mu$ that the Kalman smoother allows us to calculate the smoothed values from a state space specification.[18]

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 we implement and then present estimation results. Step by step details on the estimation procedure are provided in Appendix B.

## 5.2. Model

We adopt a three equation NK model in which a system is written for deviations of output from trend ($y_t$), inflation ($\pi_t$) and nominal interest rates ($r_t$) with three structural processes ($\zeta_{1t}$, $\zeta_{2t}$ and $\zeta_{3t}$) 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.[19] The system is given by

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

[17] 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."

[18] Appendix B provides details.

[19] We note that in our context, the theoretical mean approximation error for inflation is unbounded due to its dependence on the variance of inflation which is not finite. However, the finite sample approximation error can be small even when the theoretical mean approximation error is unbounded. Engsted et al. (2012) study the magnitude of the approximation error in a log-linearized model for stock prices, introducing extremely large deviations, in fact even explosive bubbles in stock prices, such that the prices can drift far from the linearization point. In a single equation framework, they show that despite the presence of theoretically unbounded mean approximation error, numerically there is a very small approximation error due to log-linearization as samples are $finite$. Moreover, they point out that the issue of approximation error is minimal when the sample size is small. In our context, it is also unlikely that numerically approximation errors undermine our core results since the time-series data we employ is also finite and relatively small.

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

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

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 $\mu$ to obtain our variables of interest. Prior to presenting estimation results, we discuss the potential issue of identification.

## 5.3. Identification

Given that in any DSGE empirical implementation, the identification of parameters from data is a concern, we first address this potential issue. The literature offers various approaches to estimate the parameters of a DSGE model: maximum likelihood, the method of moments, 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 RE 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 RE environments and the issue may also affect models that adopt AL (Milani, 2012).

In our context we introduce AL in an otherwise standard NK DSGE model. However, our model characterization alters the assumption on the underlying DGP 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 that 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. However, in lieu of such concerns, we conduct a direct check of whether our procedure leads to consistent estimates, and whether the filtered states are reasonably well behaved, to have confidence in the inference about the key parameters of interest, specifically the constant gain parameter.

In particular, we draw 10,000 series of each of the shocks $(\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t})$ of length 1000 from a uniform distribution. We retain a $T$ length of the errors (discarding an initial burn in) corresponding to the length of our data. For each of the 10,000 draws of shocks, we simulate our variables (output, inflation and the interest rate) using our model in (21)-(24) and Milani's (2011) estimates. We average this data and use this data for our estimation exercise. In the estimation exercise, we match the empirical and model calculated tail indices of all our variables of interest jointly. Under this specification, if we can recover Milani's estimates from our estimation procedure (which were the assumed parameters for our simulated data) we will alleviate concerns relating identification and the consistency of our estimates. We present our results from this exercise in Table 3, which confirm that the estimates are remarkably

close to Milani's estimates with extremely small standard errors.[20] These results give us confidence that the estimated parameters reported in next sections are also identified.

Table 3. Estimates (Sim. Data)

| | | All Series |
|---|---|---|
| (1) | (2) | (3) |
| | Milani | Est. |
| $\gamma_1$ | 1.4170 | 1.4194 |
| $\gamma_2$ | 0.2210 | 0.2223 |
| $\theta$ | 0.9498 | 0.9493 |
| $\kappa$ | 0.0350 | 0.0349 |
| $\tau$ | 0.2360 | 0.2373 |
| $\rho_1$ | 0.3538 | 0.3565 |
| $\rho_2$ | 0.1746 | 0.1763 |
| $\rho_3$ | N/A | 8.8E-10 |
| $\beta$ | 0.9615 | 0.9534 |
| $g$ | 0.0196 | 0.0206 |
| $\sigma_1$ | 0.7700 | 0.7751 |
| $\sigma_2$ | 0.2970 | 0.2959 |
| $\sigma_3$ | 0.2070 | 0.2022 |
| $F$ | | 9.039E-3 |

## 5.4. Results: Matching Tail Indices

Our first estimation results minimize the objective function defined as the distance between the empirical and model targets, where targets are the tail indices for all of the variables, 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 (empirical target) and $\widehat{p}^i(\mu)$, $i = y, \pi, r$ denotes the tail index from the corresponding model based counterpart (model target), given a parametrization $\mu$.

Table 4 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). These results show that apart from the constant gain parameter ($g$) the rest of the parameters are not very different from Milani (2011) which further explains small standard errors.

[20] SEs are not reported in Table 3 as they are in the order of 10E-4.

Table 4. Matching Tail Coefficients

| | | Tail Index | Tail Index |
|---|---|---|---|
| (1) | (2) | (3) | (4) |
| | Milani | Est. (Std. Err.) | Est. (Std. Err.) |
| $\gamma_1$ | 1.4170 | 1.4093 (0.0001) | 1.4093 |
| $\gamma_2$ | 0.2210 | 0.2202 (0.0001) | 0.2202 |
| $\theta$ | 0.9498 | 0.9513 (0.0001) | 0.9513 |
| $\kappa$ | 0.0350 | 0.0345 (0.0004) | 0.0345 |
| $\tau$ | 0.2360 | 0.2317 (0.0004) | 0.2317 |
| $\rho_1$ | 0.3538 | 0.3605 (0.0001) | 0.3605 |
| $\rho_2$ | 0.1746 | 0.1791 (0.0062) | 0.1791 |
| $\rho_3$ | N/A | 0.1821 (0.0434) | 0.1821 |
| $\beta$ | 0.9615 | 0.9628 (0.0012) | 0.9628 |
| $g$ | 0.0196 | 0.0513 (0.0001) | 0.0513 (0.0130) |
| $\sigma_1$ | 0.7700 | 0.7633 (0.0110) | 0.7633 |
| $\sigma_2$ | 0.2970 | 0.2903 (0.0035) | 0.2903 |
| $\sigma_3$ | 0.2070 | 0.2055 (0.0410) | 0.2055 |
| $F$ | | 3.0081E-13 | 1.151E-06 |

In addition to the direct tests conducted in Section 5.2, we now perform two additional analyses to investigate whether the estimate reported for $g$ (which is remarkably different from Milani (2011) and other literature on AL) suffers from identification issues i.e., if there are alternative values of $g$ which may also minimize our value function. 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 illustration is a way to identify if various values of $g$ can deliver convergence of value function to numerical zero. We present this illustration in Figure 1. It is clear that the minimum for the value function is only achieved at a specific estimate of $g$ as provided in Table 4.[21]

[21]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, graphical illustrations can be deceptive to the naked eye and therefore we specify the unique value of $g$ and the corresponding converged value of the objective function at the top of the graph.

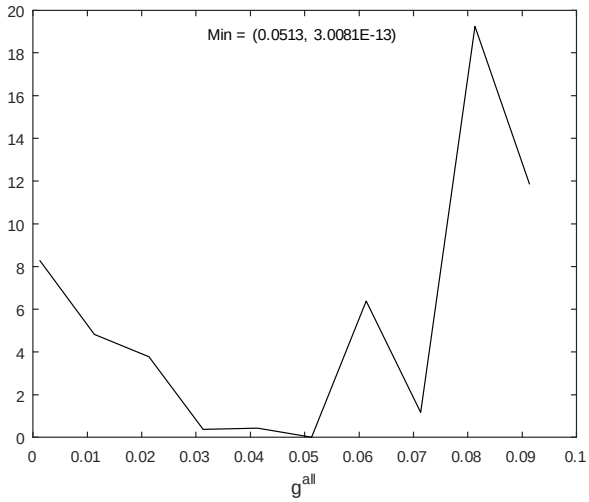


Figure 1. The Objective Function ($F$: Equation 42) vs. $g$.

In our second analysis, we only estimate the constant gain ($g$) and fix all the values of the parameters in $\mu$ at the values given in Table 4, column 3. We choose these values for two reasons. First, the estimates are not very different from Milani (2011); second, it provides us with another check to investigate if we can recover the value of $g$ from Table 4 column 3. Moreover, this analysis allows us to address the issue of estimating more parameters while minimizing the distance between the empirical and model targets in our last estimation. We present the results of the estimated $g$ and the corresponding standard errors in Table 4, column 4. We recover the same value of $g$ as reported in column 3.

Given our results show a larger than usual estimated value for the gain parameter, we explore an additional concern. 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 RE and thus model variables should have larger tail indices indicating proximity to Normality.

Recall that the LRMN representation of our system of interest is given by equations (17)-(20). Our first step in conducting the simulations is to draw shocks ($\varepsilon_t$) from a known distribution. We choose a uniform distribution with support from 0 to 1 and draw 5000 series of length 1000. For each $\varepsilon_t$ series of length $T = 1000$, we simulate the above system using as a baseline the parameter estimates from Table 4, column 3. 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 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 2-4. They 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.

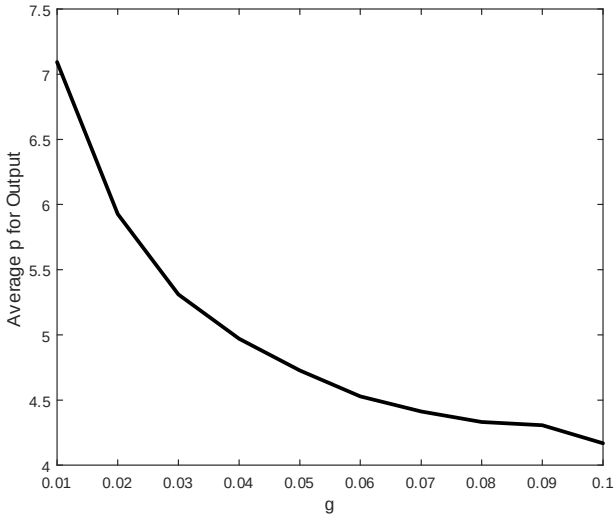


Figure 2. Simulated Output Tail Indices vs. $g$.

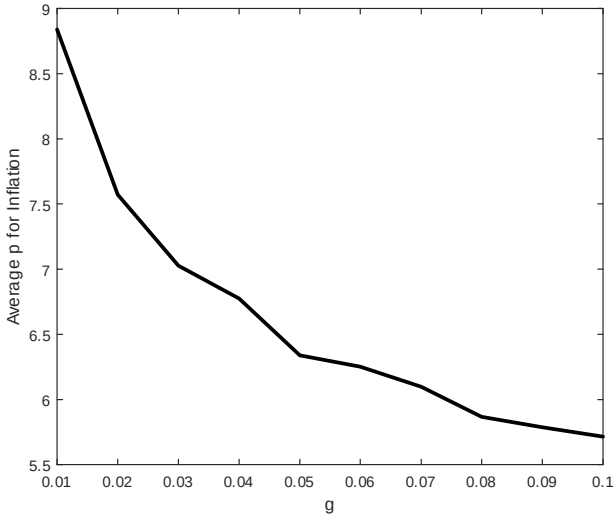


Figure 3. Simulated Inflation Tail Indices vs. $g$.

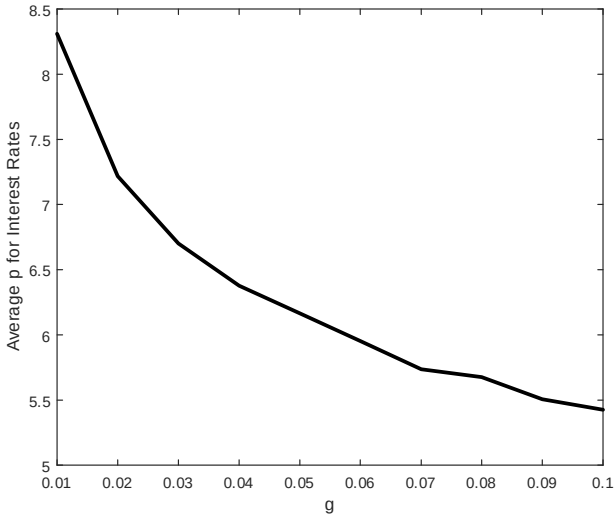


Figure 4. Simulated Interest Rate Tail Indices vs. $g$.

## 5.5. Results: Matching Tail Indices and Moments

The analyses in the last estimation exercises minimized the distance of our objective function by matching the tail indices only. However, this raises an alternative concern that while our model is successful in matching fat tails it may erode model performance on other dimensions. Moreover, matching only a few targets reduces the number of free parameters one may be able to estimate. To tackle these concerns, we expand the targets we match in our next two estimation exercises while estimating more parameters.

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:

$$n = \widehat{p} - 1. \tag{43}$$

Higher moments such that $n > \widehat{p} - 1$ are infinite and therefore do not exist. Recall from Table 2 that the tail indices for quarterly HP-filtered output, inflation and interest rates are as follows: $p^y = 3.6395$ , $p^\pi = 2.4898$ and $p^r = 4.8613$. As a result the number of moments that are finite for output, inflation and interest rate, respectively are $n^y = 2$, $n^\pi = 1$ 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 $(m_1^\pi)$ exist for inflation and the first three moments i.e., mean $(m_1^r)$, standard deviation $(m_2^r)$ and skewness $(m_3^r)$ exist for the interest rate. Using this information we re-estimate our parameters by using an objective function that consists only of the the value of finite moments, determined by the tail indices in the data. In the second exercise we include both the value of moments and the value of tail indices for all the variables jointly.

Corresponding objective functions for these exercises therefore are as follows:

$$\min_{\mu} F = \left[ \begin{pmatrix} m_1^y \\ m_2^y \\ m_1^\pi \\ m_1^r \\ m_2^r \\ m_3^r \end{pmatrix} - \begin{pmatrix} \widehat{m}_1^y(\mu) \\ \widehat{m}_2^y(\mu) \\ \widehat{m}_1^\pi(\mu) \\ \widehat{m}_1^r(\mu) \\ \widehat{m}_2^r(\mu) \\ \widehat{m}_3^r(\mu) \end{pmatrix} \right]' \left[ \begin{pmatrix} m_1^y \\ m_2^y \\ m_1^\pi \\ m_1^r \\ m_2^r \\ m_3^r \end{pmatrix} - \begin{pmatrix} \widehat{m}_1^y(\mu) \\ \widehat{m}_2^y(\mu) \\ \widehat{m}_1^\pi(\mu) \\ \widehat{m}_1^r(\mu) \\ \widehat{m}_2^r(\mu) \\ \widehat{m}_3^r(\mu) \end{pmatrix} \right]. \tag{44}$$

$$\min_{\mu} F = \left[ \begin{pmatrix} p^y \\ p^\pi \\ p^r \\ m_1^y \\ m_2^y \\ m_1^\pi \\ m_1^r \\ m_2^r \\ m_3^r \end{pmatrix} - \begin{pmatrix} \widehat{p}^y(\mu) \\ \widehat{p}^\pi(\mu) \\ \widehat{p}^r(\mu) \\ \widehat{m}_1^y(\mu) \\ \widehat{m}_2^y(\mu) \\ \widehat{m}_1^\pi(\mu) \\ \widehat{m}_1^r(\mu) \\ \widehat{m}_2^r(\mu) \\ \widehat{m}_3^r(\mu) \end{pmatrix} \right]' \left[ \begin{pmatrix} p^y \\ p^\pi \\ p^r \\ m_1^y \\ m_2^y \\ m_1^\pi \\ m_1^r \\ m_2^r \\ m_3^r \end{pmatrix} - \begin{pmatrix} \widehat{p}^y(\mu) \\ \widehat{p}^\pi(\mu) \\ \widehat{p}^r(\mu) \\ \widehat{m}_1^y(\mu) \\ \widehat{m}_2^y(\mu) \\ \widehat{m}_1^\pi(\mu) \\ \widehat{m}_1^r(\mu) \\ \widehat{m}_2^r(\mu) \\ \widehat{m}_3^r(\mu) \end{pmatrix} \right]. \tag{45}$$

While minimizing these additional objective functions, we estimate an expanded set of parameters which allows us to investigate not only how the constant gain ($g$) alters but also how the transmission mechanism changes under our new objective functions. In particular for the first objective function, we estimate the following parameters $\mu = [\gamma_1, \gamma_2, \theta, \kappa, \tau, g]$ and for the second exercise our estimated parameters include $\mu = [\gamma_1, \gamma_2, \theta, \kappa, \tau, \rho_1, \rho_2, \rho_3, g]$. All other parameters not in $\mu$ are fixed at Milani's posterior mean. Table 5 presents the main results for these estimation exercises: columns 2 and 3 present Milani's estimates and the 95% confidence intervals while columns 4 and 5 respectively present our estimates and standard errors for our parameters under objective function 44 and 45.

We draw three inferences. First, the estimated $g$ is consistently large and echoes our previous results while minimizing our new objective functions. Second, we observe some deviation from Milani's (2011) posterior mean values reported in column 3, however, almost all of our estimates in column 4 and 5 (barring $\theta$ and $\rho_1$ which deviate only slightly), lie within Milani's 95% credible intervals reported in column 3. Third, since a priori it is not clear from a theoretical perspective how changing the objective function in our estimation exercise (by sequentially increasing the number of moments we match) relates to the underlying linear recursion with multiplicative noise (LRMN) that governs the adaptive learning environment, this exercise allows us to investigate the sensitivity of $g$ to changing objective functions.[22]

[22] Theoretically we know that (i) as the probability mass above one of the multiplicative noise term ($P(\Phi_t > 1)$) increases then the learning environment would deliver increased volatility and dispersion (ii) we also know that the constant gain that tries to match volatility and dispersion is related to that mass as shown in the simulation based comparative statics (as the gain increases the tail indices fall). It could be that on the one hand as the econometric objective function $F$ is made richer that indeed a higher $g$ is needed as more information is being used to estimate parameters. Alternatively it could be that the increased amount of information is redundant as all the underlying unknown DGP needed was already contained in the tail index

Our results indicate that quantitatively our estimated of $g$ is very similar across various objective functions.

Table 5. Matching Tail Coefficients or/and Moments

| | | | Moments | Tail and Moments |
|---|---|---|---|---|
| (1) | (2) | (3) | (4) | (5) |
| | Milani | 95% Credible Interval | Est. (Std. Err.) | Est. (Std. Err.) |
| $\gamma_1$ | 1.4170 | [0.97 - 1.86] | 1.5975 (0.0001) | 1.5177 (0.0012) |
| $\gamma_2$ | 0.2210 | [0.06 - 0.43] | 0.3513 (0.0012) | 0.1212 (0.0003) |
| $\theta$ | 0.9498 | [0.91 - 0.98] | 0.9305 (0.0002) | 0.8527 (0.0048) |
| $\kappa$ | 0.0350 | [0.019 - 0.053] | 0.0420 (0.0024) | 0.0231 (0.0044) |
| $\tau$ | 0.2360 | [0.03 - 0.55] | 0.0359 (0.0002) | 0.3027 (0.0002) |
| $\rho_1$ | 0.3538 | [0.19 - 0.50] | 0.3538 | 0.5466 (0.0336) |
| $\rho_2$ | 0.1746 | [0.04 - 0.31] | 0.1746 | 0.2662 (0.0001) |
| $\rho_3$ | N/A | N/A | 0.0000 | 0.0034 (0.0001) |
| $\beta$ | 0.9615 | N/A | 0.9615 | 0.9615 |
| $g$ | 0.0196 | [0.015 - 0.025] | 0.0571 (0.0042) | 0.0572 (0.0032) |
| $\sigma_1$ | 0.7700 | [0.69 - 0.86] | 0.7700 | 0.7700 |
| $\sigma_2$ | 0.2970 | [0.27 - 0.33] | 0.2970 | 0.2970 |
| $\sigma_3$ | 0.2070 | [0.19 - 0.23] | 0.2070 | 0.2070 |
| $F$ | | | 2.94E-08 | 8.97E-04 |

To again investigate any potential identification issues that may be of a concern if various values of a particular parameter can deliver convergence of the value function to numerical zero, we present Figures 5 and 6 respectively for our objective functions described above. These figures further moderate concerns regarding identification issues.

to begin with. This is why comparison across the estimates of $g$ based on different objective functions is an empirical question, described in this section.

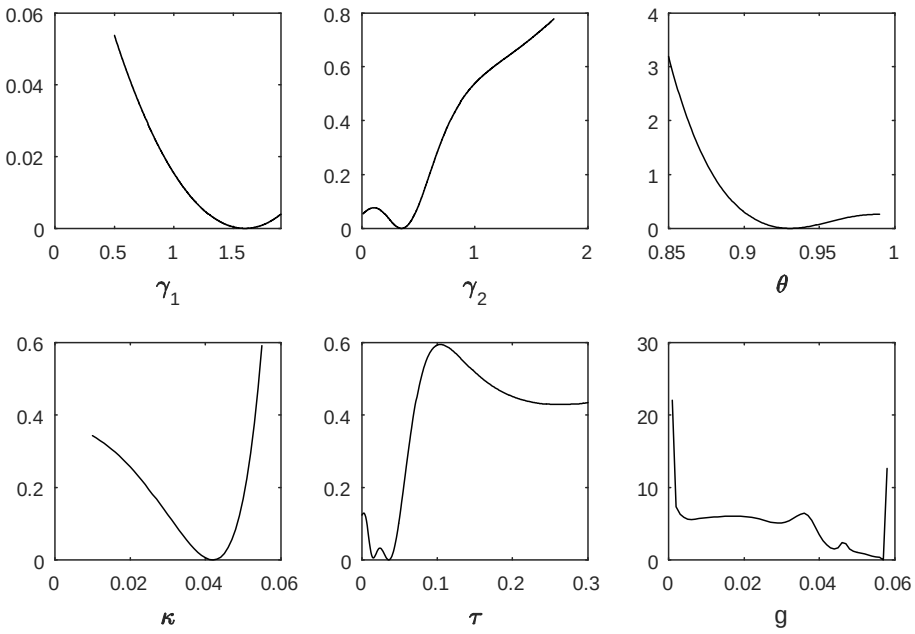


Figure 5. The Objective Function ($F$: Equation 49) vs. $\mu = [\gamma_1, \gamma_2, \theta, \kappa, \tau, g]$.

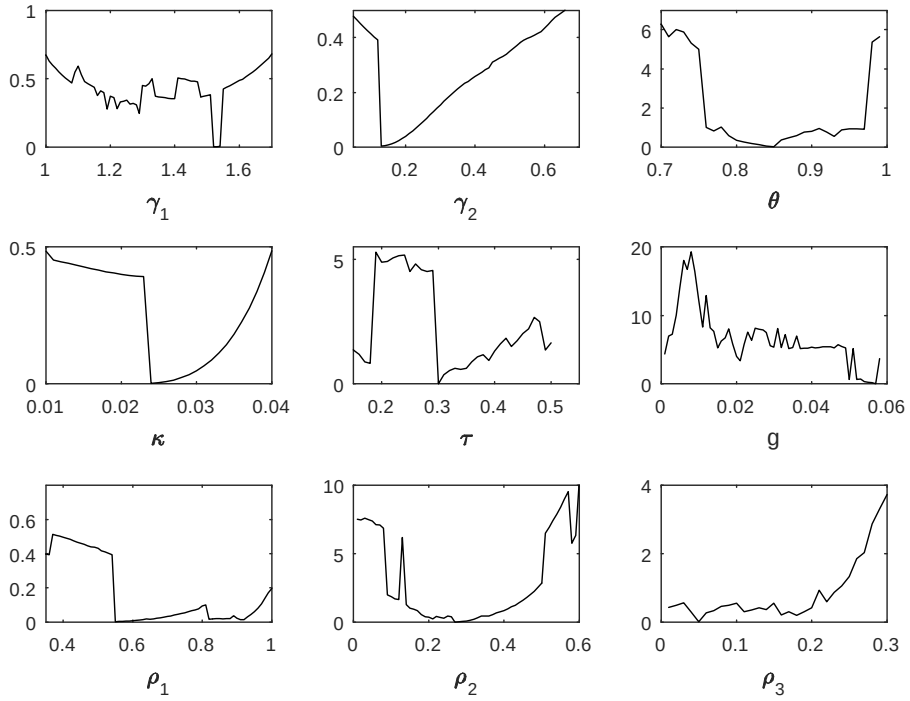


Figure 6. The Objective Function ($F$: Equation 50) vs. $\mu = [\gamma_1, \gamma_2, \theta, \kappa, \tau, g, \rho_1, \rho_2, \rho_3]$.

The foremost conclusion from Table 5 and the corresponding figures is that our estimated constant learning gain parameter ($g$) is bigger than the estimate usually documented in the

existing literature (between 0.01-0.03) and the 95% credible interval reported by Milani. 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 $g$.

### 5.6. Summary

Our model, estimation results and the analyses present an endogenous channel via SGCG learning that delivers fat tails for variables in an otherwise standard model. We show that without altering the nature of structural shocks and the transmission channel, a DSGE model under AL endogenously delivers model dynamics that better replicate observed fat tails without eroding the model's performance on other dimensions. In particular, we see that the baseline estimation, the estimation that accounts for potential identification issues, and the expanded estimation exercises consistently 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 results sustain when the shock size is fixed at the values as used in the literature. Therefore larger than usual shocks do not seem to be driving the dynamics to deliver fat tails in the data under the LRMN assumption.

Moreover, in our expanded estimation exercises where our targets are both tails and moments, we also show a high gain in our model is responsible in generating fat tails without sacrificing the fit of the model regarding the finite moments of the distribution of output, inflation and interest rate. These results highlight that a minimal role may remain for alternate structural parameters that govern transmission mechanisms, in accounting for fat tails in macroeconomic data. We further show in the next section that a model without the constant learning mechanism is less successful in matching the statistical regularity of fatter tails in data. This evidence reassures us that absent learning mechanisms, other transmission mechanisms may not be enough to account for observed fat tails. In summary, we conclude that our core result of a higher constant gain is robust to the various specifications provided in the prior sections.

## 6. Simulations

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 $\varepsilon_t$ and retain a $T$ length of errors (discarding an initial burn in) corresponding to the length of our data. For each of the 1000 series, we simulate the model for $y_t$, $\pi_t$ 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, $\varepsilon_t$ 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 $\varepsilon_t$ 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 simulations, we again close the model under RE but instead replace the Normal distribution of $\varepsilon_t$ with a Student's-t non-Gaussian process that has a priori fatter tails. 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 7 and Figure 8. 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 9 and Figure 10.

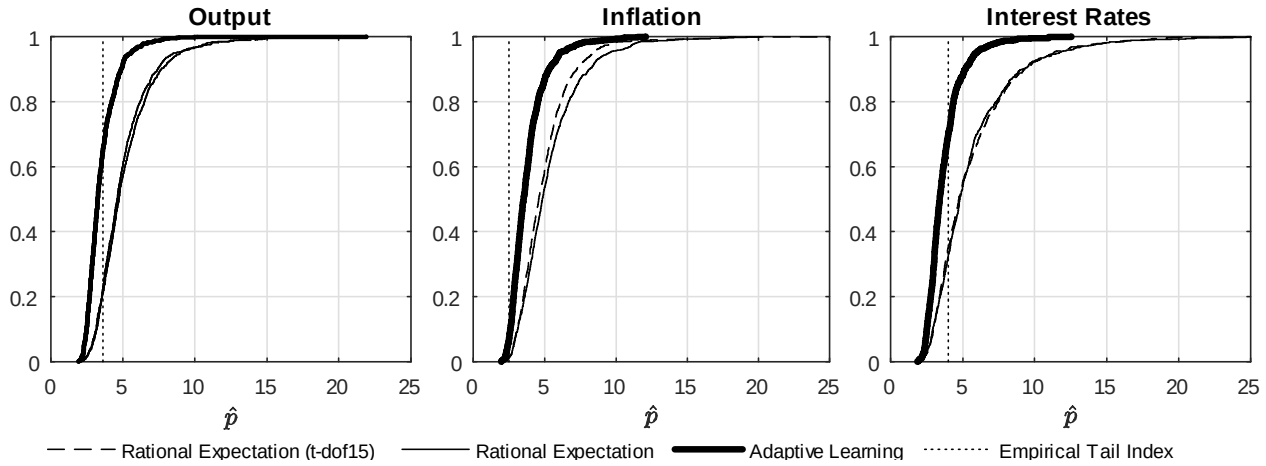


Figure 7. Simulated CDFs.

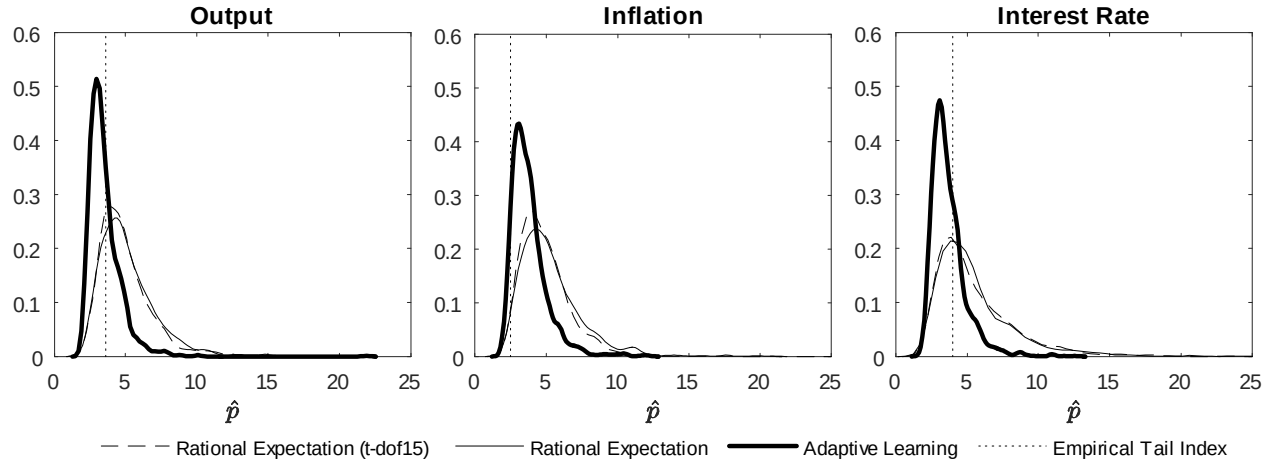


Figure 8. Simulated PDFs.

Figure 7 (Figure 9) and Figure 8 (Figure 10), 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 7. First, the probability of $p$ taking a value closer to the $p$ estimated from the data, is always higher under the AL model than under other alternatives. Second, for each of output, inflation and the interest rate, the estimates of $p$ from other alternatives first order stochastically dominate the estimates of $p$ under AL. 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 AL. Figure 8 illustrates the same results using the probability distribution function and shows that the average $p$ under AL is closer to the $p$ in the data. Furthermore, in Table 6. (Table 7.) Kolmogorov-Smirnov test statistics at 1 percent significance level reject the null hypothesis that the two samples (from our model versus other alternatives) are drawn from the same distribution.

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 |
| $\pi_t$ | 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 |
| $\pi_t$ | 0.0000 | 1 | 0.0036 | 1 | 0.0000 | 1 |
| $r_t$ | 0.0000 | 1 | 0.0000 | 1 | 0.0000 | 1 |

Figures 9 and 10 show similar results as previously 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 the 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 use of data.

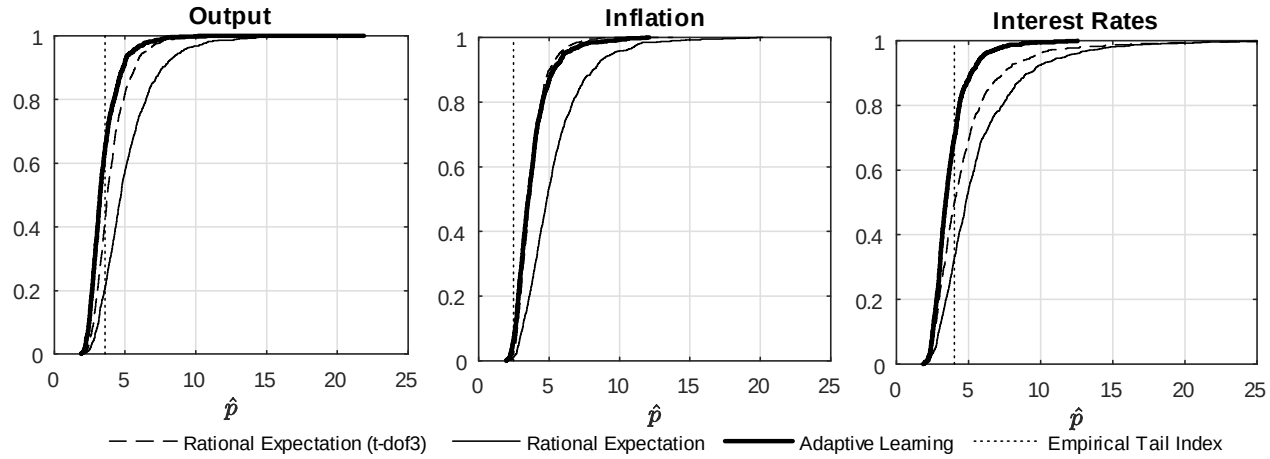


Figure 9. Simulated CDFs.

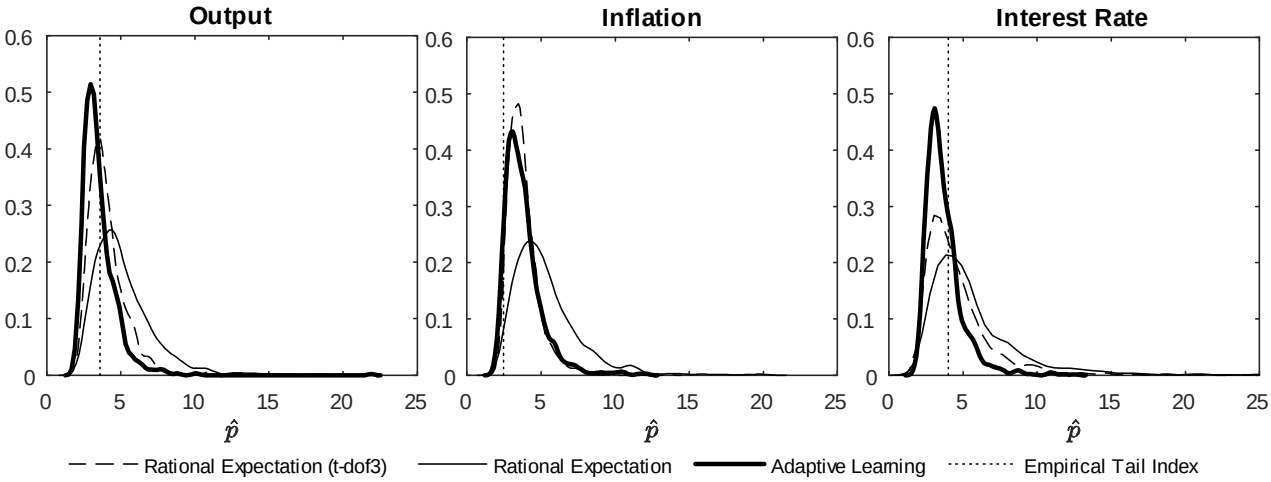


Figure 10. Simulated PDFs.

# 7. Conclusion

In accounting for large cyclical fluctuations of macroeconomic series two approaches are available. The first assumes non-Normal distributions for innovations to structural shocks and has seen some success in being empirically validated. The second proposes an endogenous mechanism, adaptive learning, which delivers a model representation that entails multiplicative and additive noise. We evaluate 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. This finding is similar in spirit to the debate implicit in Sims (2001) and Cogley and Sargent (2001, 2005) who were concerned 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 and additive noise.

Next, our simulation exercises indicate a monotonically decreasing relationship between the gain and the possibility of fat tails in macroeconomic aggregates. Moreover they suggest that AL outperforms fat tailed distributional assumptions on innovations to exogenous structural shocks in being able to match stylized facts along the fat tail dimension. Taken together, our empirical and simulation exercises suggest that AL representations can lead to regularly occurring large fluctuations in macroeconomic systems, and match reality.

It is important to point out that, in Hanson's (2014) terminology, our work is contributing to the "outside econometrician" perspective since we aim to reconcile fat tails observed in the data with the model and therefore simplify the framework by using a representative agent set-up. However, it is equally interesting to explore how an "inside econometrician" perspective captured by a heterogenous agent framework would allow in reconciling the fat tails in data with the model. Recent empirical work (see e.g., Carroll, 2003; Mankiw et al., 2003; Branch, 2004) documents the disagreements and heterogeniety in the beliefs of agents in the survey data. However, theoretically, only a scant literature allows different fractions of agents to form expectations that correspond to either rational and/or boundedly rational/adapative expectations (e.g., Levine (2010), Branch and McGough (2016)). In our context, it is an open question as to how heterogeneous beliefs of agents impact reduced form relations governing the evolution of the economy, which Branch and McGough (2016) also point out is an open issue.[23]

[23] To model an "inside econometrician perspective" whereby agents are modeled in more detail, one would move from the representative agent model (as is presented in this paper) towards an heterogenous agent framework. It is interesting to note that in such a setting an agent who defects from using lets say a decreasing gain in favor of constant gain (while majority of agents use decreasing gain), the defector would be worse off. This is in constrast with the "outside econometrician perspective" presented in this paper whereby the goal is to match the volatility as observed in the data which is better captured by the constant gain versus decreasing gain. We thank a referee for pointing out the intuition of various gain rules in the heterogenous setting instead of representative agent setting as in our set up.

# 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., 2004. The theory of rationally heterogeneous expectations: evidence from survey data on inflation expectations. The Economic Journal, 114(497), pp.592-621.

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

[14] Branch, W.A. and McGough, B., 2009. A New Keynesian model with heterogeneous expectations. Journal of Economic Dynamics and Control, 33(5), pp.1036-1051.

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

[16] Carroll, C.D., 2003. Macroeconomic expectations of households and professional forecasters. The Quarterly Journal of Economics, 118(1), pp.269-298.

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

[18] 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.

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

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

[21] 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.

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

[23] 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.

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

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

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

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

[28] Engsted, T., Pedersen, T.Q. and Tanggaard, C., 2012. The log-linear return approximation, bubbles, and predictability. Journal of Financial and Quantitative Analysis, 47(3), pp.643-665.

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

[30] Evans, G.W., Honkapohja, S. and Williams, N., 2010. Generalized stochastic gradient learning. International Economic Review, 51(1), pp.237-262.

[31] 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.

[32] 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.

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

[34] 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.

[35] Hansen, L.P., 2014. Nobel lecture: Uncertainty outside and inside economic models. Journal of Political Economy, 122(5), pp.945-987.

[36] 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).

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

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

[39] 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.

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

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

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

[43] Mankiw, N.G., Reis, R. and Wolfers, J., 2003. Disagreement about inflation expectations. NBER macroeconomics annual, 18, pp.209-248.

[44] 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.

[45] 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.

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

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

[48] 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.

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

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

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

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

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

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

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

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

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

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

[59] 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.

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

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

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

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

[64] 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.

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

[66] 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.

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

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

# Appendix A: Goodness of Fit Test

In this Appendix we detail the method used in conducting the goodness of fit test for our data. This procedure allows us to formally test the hypothesis of whether the distribution of the data is power law and consequently if the fat tail indices are significant and meaningful.

1. Fit the empirical data to the power law model and estimate $\widehat{p}$ and the lower $Y_{\min}$ using the methods described in Section 4.1 and calculate the Kolmogorov-Smirnov (KS) test statistic for this fit. In particular, KS statistics is simply the maximum distance between the CDFs of the data and the fitted model:

$$D = \max_{y \geq Y_{\min}} |S(y) - P(y)|$$

   where $S(y)$ is the CDF of the data for the observations with the value at least $Y_{\min}$ and $P(y)$ is the CDF for the power law model that best fits the data in the region $y_i > Y_{\min}$

2. To generate a large number of synthetic power-law distributed data sets with the same $\widehat{p}$ and $Y_{\min}$ (as estimated in (1)) as well as same length as the empirical data ($T$), we proceed in two steps. Denote by $\eta_{tail}$ the number of obsevations above $Y_{\min}$. First, to construct the synthetic data above $Y_{\min}$ ($y_i \geq Y_{\min}$), we generate with a probability $\eta_{tail}/T$ a random number ($y_i$) drawn from a power-law distribution (which shares the same $\widehat{p}$ and $Y_{\min}$ as our empirical data but the observation are from the fitted power law). However, to construct the synthetic data below $Y_{\min}$ ($y_i < Y_{\min}$) we generate, with a probability $1 - \eta_{tail}/T$, a random number $y_i$ such that we select one element uniformly at random from among the elements of our empirical dataset. We repeat this for all the $T$ observations and this provide us with a comparable synthetic datasets which follow a power law above $Y_{\min}$ but has the same non-power law distribution as our empirical data. We construct many of such synthetic datasets, say 10,000.

3. Fit each synthetic data (constructed in step (2)) to its power law model and calculate the KS statistics. Then compare the KS statistics for each synthetic data with that of our empirical dataset from step 1 and count what fraction of the time the resulting statistic is larger than the value for the empirical data. This fraction is our $p$-value.

4. We repeat the process for each of the 10,000 synthetic dataset so that our $p$-value is accurate upto 3 decimal places. To do so we choose a very conservative $p$-value of 0.1 as a cut-off point such that $p < 0.1$ rules out the power law and the associated $\widehat{p}'s$. $p = 0.1$ means that 1 in 10 cases, we would merely by chance get data that agree as poorly with the model as the data we have.

# Appendix B: Structural Model, Kalman Filter, Standard Errors

Consider a linear New Keynesian (NK) model featuring variables in deviation from steady state so there are no constant terms,

$$\begin{aligned}
0 &= -y_t - \sigma i_t + \sigma\zeta_{1t} + E_t(y_{t+1}) + \sigma E_t(\pi_{t+1}), && (46)\\
0 &= \kappa y_t - \pi_t + \zeta_{2t} + \beta E_t(\pi_{t+1}), && (47)\\
0 &= (1-\theta)\gamma_2 y_{t-1} + (1-\theta)\gamma_1\pi_{t-1} + \theta i_{t-1} + (1-\theta)\zeta_{1t} + \zeta_{3t} - i_t, && (48)\\
\zeta_{1t} &= \rho_1\zeta_{1t-1} + \varepsilon_{1t},\ \varepsilon_{1t} \sim iid(0,\sigma_1^2), && (49)\\
\zeta_{2t} &= \rho_2\zeta_{2t-1} + \varepsilon_{2t},\ \varepsilon_{2t} \sim iid(0,\sigma_2^2), && (50)\\
\zeta_{3t} &= \rho_3\zeta_{3t-1} + \varepsilon_{3t},\ \varepsilon_{3t} \sim iid(0,\sigma_3^2). && (51)
\end{aligned}$$

The first equation is the IS curve (relating output ($y$) with inflation ($\pi$) and nominal interest rates ($i$)), followed by a Phillips curve and a Taylor rule. All three of these structural equations feature autoregressive shocks ($\zeta_i$) with $iid$ innovations ($\varepsilon_i$) and the six equations can be written in the forms needed under rational expectations (RE) and adaptive learning (AL).

In this Appendix we first describe how RE delivers a fixed coefficient VAR representation of a model that then is adopted as the state equation in a state space system. Next, we show that under SGCG learning the model representation is a time varying coefficient VAR representation that too can be adopted as the state equation in a state space system. Having obtained the two fundamental representations we then show that the Kalman filter and smoother recursions apply and argue how exactly distributional assumptions come into play when specifying likelihood functions, versus the minimum distance empirical exercise we conducted. In the final section we show how the model is written as a linear recursion with multiplicative noise (LRMN) that clearly demonstrates how innovations, however small in a linear model, can accumulate under constant gain learning such that the tail of the stationary distribution of model variables can be fat, implying a higher probability that the model exhibits what we term as a "large fluctuation" from trend.

## 7.1. Model Representation: Rational Expectations

The NK model specified above can be written as

$$\Gamma_0(\theta)X_t = \Gamma_1(\theta)X_{t-1} + \Psi(\theta)\varepsilon_t + \Pi(\theta)\eta_t. \quad (52)$$

Solving the above under RE following DeJong and Dave (2011)'s exposition of Sims (2001), a unique stable solution takes a familiar fixed coefficient VAR form:

$$X_t = F(\theta)X_{t-1} + G(\theta)\varepsilon_t, \quad (53)$$

where as usual the matrices $F$ and $G$ contain as elements values (or combinations) of $\theta$, the deep parameters. We note that in order to observe "large" swings in $X_t$ or elements thereof,

one can either assume large values for $\varepsilon_t$ directly or assume that $\varepsilon_t$ is drawn from a fat tailed distribution. Therefore, under the RE fixed coefficient VAR representation of any model, exogenous assumptions on $\varepsilon_t$ deliver "large" swings in $X_t$ or elements thereof.

The above equation is said to be the state equation of a time invariant state-space system when coupled with an observer equation, as follows

$$\begin{aligned} X_t &= F(\theta)X_{t-1} + G(\theta)\varepsilon_t, & (54)\\ Y_t &= H'X_t, & (55) \end{aligned}$$

where $Y_t$ denotes HP filtered data of dimension $K$. Given distributional assumptions on the structural innovations $\varepsilon_t$ the above system is fed into a Kalman filter yielding a likelihood function $L(\theta)$ that is utilized in either a frequentist maximum likelihood approach to estimation or coupled with prior distributions $p(\theta)$ in order to calculate Bayesian posteriors. Having obtained numerical values (estimates) for $\theta$, a Kalman smoother can then be employed to recursively calculate fitted variables $(\widehat{X}_t)$. We describe this procedure below after we describe next the representation of a model under adaptive learning.

## 7.2. Model Representation: Adaptive Learning

Under AL a solution of the model assumes first that the representative agent knows all data up until $t-1$ so that the model is given by

$$X_t = A(\theta)E_{t-1}(X_{t+1}) + B(\theta)U_t. \quad (56)$$

Next, a perceived law of motion (PLM) on the part of agents is conjectured as a time varying analog to the RE solution:

$$X_t = b_t X_{t-1} + \xi_t \quad (57)$$

and given the assumption that only data up until $t-1$ are known

$$\begin{aligned} X_{t+1} &= b_{t+1}X_t + \xi_{t+1} & (58)\\ E_{t-1}(X_{t+1}) &= b_{t-1}E_{t-1}(X_t) & (59)\\ E_{t-1}(X_t) &= b_{t-1}E_{t-1}(X_{t-1}) = b_{t-1}X_{t-1} & (60)\\ &\rightarrow \; E_{t-1}(X_{t+1}) = b_{t-1}^2 X_{t-1} & (61) \end{aligned}$$

which is then inserted into the original model to obtain the actual law of motion (ALM) as

$$\begin{aligned} X_t &= A(\theta)b_{t-1}^2 X_{t-1} + B(\theta)U_t, & (62)\\ U_t &= PU_{t-1} + \varepsilon_t & (63) \end{aligned}$$

where the evolution of $b_t$ is governed by the constant gain learning rule

$$b_t = b_{t-1} + gX_{t-1}(X_t' - X_{t-1}'b_{t-1}),\ b_1 = (X_1X_1')^{-1}. \quad (64)$$

Initial beliefs are fixed at $b_1 = (X_1X_1')^{-1}$. We note our focus on a constant gain learning rule versus a decreasing gain learning rule that would replace $g$ with a decreasing function

of time (e.g. $t^{-1}$). A constant gain rule allows us to incorporate the notion that only a fixed past is used in forming expectations. The adoption of decreasing gains is an interesting comparison to SGCG learning but would not allow us to make a comparison of this model representation with the LRMN model representation (discussed below). This is important as it is the congruency between the two model representations under constant gain that allows for the intuition of large movements in model variables due to a particular accumulation of otherwise small shocks.

Next we need to write the overall state equation, as follows. Let $Z_t = [X_t \quad U_t]'$ then

$$\begin{bmatrix} X_t \\ U_t \end{bmatrix} = \begin{bmatrix} I & -\kappa \\ {\scriptstyle m\times m} & {\scriptstyle m\times p} \\ 0 & I \\ {\scriptstyle p\times m} & {\scriptstyle p\times p} \end{bmatrix}^{-1} \begin{bmatrix} \beta b_{t-1}^2 & 0 \\ {\scriptstyle m\times m} & {\scriptstyle m\times p} \\ 0 & K \\ {\scriptstyle p\times m} & {\scriptstyle p\times p} \end{bmatrix} \begin{bmatrix} X_{t-1} \\ U_{t-1} \end{bmatrix} + \begin{bmatrix} I & -\kappa \\ {\scriptstyle m\times m} & {\scriptstyle m\times p} \\ 0 & I \\ {\scriptstyle p\times m} & {\scriptstyle p\times p} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ {\scriptstyle m\times p} \\ I \\ {\scriptstyle p\times p} \end{bmatrix} \varepsilon_t, \tag{65}$$

$$Z_t = A_t Z_{t-1} + B\varepsilon_t. \tag{66}$$

Introducing again the observation vector $Y_t$ allows us to specify a time-varying state-space model

$$Z_t = A_t Z_{t-1} + B\varepsilon_t, \tag{67}$$

$$Y_t = H' Z_t. \tag{68}$$

Note that this state space too can be Kalman filtered and smoothed, just like in the rational expectations case, to obtain smoothed values of $Z_t$, described next.

## 7.3. Kalman Filtering and Smoothing

Using stand alone notation in this sub-section, consider a fixed coefficient (RE) state space system

$$X_t = FX_{t-1} + e_t, \; E(e_t e_t') = \Psi, \tag{69}$$

$$Y_t = H' X_t, \tag{70}$$

where $e_t = G\varepsilon_t$ relative to previous notation. then following DeJong and Dave (2011) and Lutkepohl (2007, pp. 626-633) the Kalman recursions can be defined by first defining

$$X_{t|s} = E(X_t|Y_1, ..., Y_s), \; \Sigma_X(t|s) = COV(X_t|Y_1, ..., Y_s), \tag{71}$$

$$Y_{t|s} = E(Y_t|Y_1, ..., Y_s), \; \Sigma_Y(t|s) = COV(Y_t|Y_1, ..., Y_s). \tag{72}$$

(We note that this sub-section applies equally to variable coefficient (AL) state space systems using a change of notation). Given definition of the above expectations, the prediction equations set for $1 \leq t \leq T$ of the filter is given by

$$X_{t|t-1} = FX_{t-1|t-1}, \tag{73}$$

$$\Sigma_Y(t|t-1) = F\Sigma_Y(t-1|t-1)F' + \Psi, \tag{74}$$

$$Y_{t|t-1} = HX_{t|t-1}, \tag{75}$$

$$\Sigma_Y(t|t-1) = H\Sigma_X(t|t-1)H', \tag{76}$$

the correction equations set for $1 \leq t \leq T$ of the filter is given by

$$\begin{aligned} X_{t|t} &= X_{t|t-1} + P_t(Y_t - Y_{t|t-1}), && (77) \\ \Sigma_X(t|t) &= \Sigma_X(t|t-1) - P_t\Sigma_Y(t|t-1)P_t', && (78) \\ P_t &= \Sigma_X(t|t-1)H'\Sigma_Y(t|t-1)^{-1} \text{ [Kalman Gain]}, && (79) \end{aligned}$$

the forecasting equations set for $t > T$ of the filter is given by

$$\begin{aligned} X_{t|T} &= FX_{t-1|T}, && (80) \\ \Sigma_X(t|T) &= F\Sigma_X(t-1|T)F' + \Psi, && (81) \\ Y_{t|T} &= HX_{t|T}, && (82) \\ \Sigma_Y(t|T) &= H\Sigma_X(t|T)H', && (83) \end{aligned}$$

and the smoothing set of equations for $t < T$ of the Kalman smoother is given by

$$\begin{aligned} \widehat{X}_t &\equiv X_{t|T} = X_{t|t} + S_t(X_{t+1|T} - X_{t+1|t}), && (84) \\ \Sigma_X(t|T) &= \Sigma_X(t|t) - S_t[\Sigma_X(t+1|t) - \Sigma_X(t+1|T)]S_t', && (85) \\ S_t &= \Sigma_X(t|t)F'\Sigma_X(t+1|t)^{-1}. && (86) \end{aligned}$$

Given the above recursions, if a likelihood approach (frequentist or Bayesian) is to be employed to estimate the deep parameters $\theta$ then that likelihood function can be built using as input some of the above recursions as

$$\ln L(\theta|Y_1, ..., Y_T) = -\frac{KT}{2}\ln 2\pi - \frac{1}{2}\sum_{t=1}^{T}\ln|\Sigma_Y(t|t-1)| - \frac{1}{2}\sum_{t=1}^{T}(Y_t - Y_{t|t-1})'\Sigma_Y(t|t-1)^{-1}(Y_t - Y_{t|t-1}). \quad (87)$$

We note the following about the recursions above following DeJong and Dave (2011) and Lutkepohl (2007). First, the above recursions are specified for a fixed coefficient state space system

$$\begin{aligned} X_t &= FX_{t-1} + e_t, \; E(e_te_t') = \Psi, && (88) \\ Y_t &= H'X_t. && (89) \end{aligned}$$

However, the same recursions can be specified, given suitable notation, for a variable coefficients system

$$\begin{aligned} X_t &= F_tX_{t-1} + e_t, \; E(e_te_t') = \Psi_t, && (90) \\ Y_t &= H'X_t, && (91) \end{aligned}$$

where the subscript $t$ on $F_t$ and $\Psi_t$ denote time variation. Second, Lutkepohl (2007. p. 626) writes that "If the normality assumption is dropped, the recursions given below can still be justified." Similarly, the Kalman filter and smoother recursions given above can be used to construct smoothed states $\widehat{X}_t$ from either a fixed coefficients or variable coefficients state space system even if the white noise processes $e_t$ are not Gaussian. That RE delivers

a fixed coefficient state space system and constant gain AL delivers a variable coefficients state space representation of a model are thus both described above.

## 7.4. Model Representation: LRMN

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

$$X_t = \hat{\Phi}_t U_t \tag{92}$$

which is still only a slight departure from the above technique since a VAR can always be written in MA form. Next, assume that $E_{t-1}(X_{t+1}) = \hat{\Phi}_{t-1} U_t$ which yields the ALM

$$X_t = A(\theta)\hat{\Phi}_{t-1} U_t + B(\theta) U_t. \tag{93}$$

The evolution of $\hat{\Phi}_t$ is governed by the constant gain learning rule

$$\hat{\Phi}_t = \hat{\Phi}_{t-1} + g U_{t-1}(X_t - \hat{\Phi}_{t-1} U_{t-1}) \tag{94}$$

and inserting the ALM in place of $X_t$ into this rule yields

$$\begin{aligned}
\hat{\Phi}_t &= \hat{\Phi}_{t-1} + g U_{t-1}(A(\theta)\hat{\Phi}_{t-1} U_t + B(\theta) U_t - \hat{\Phi}_{t-1} U_{t-1}), && (95)\\
&= \hat{\Phi}_{t-1} + g A(\theta) U_{t-1}\hat{\Phi}_{t-1} U_t + g B(\theta) U_{t-1} U_t - g U_{t-1}\hat{\Phi}_{t-1} U_{t-1}, && (96)\\
&= \left[I + g A(\theta) U_{t-1} U_t - g U_{t-1} U_{t-1}\right]\hat{\Phi}_{t-1} + g B(\theta) U_{t-1} U_t. && (97)
\end{aligned}$$

The learning model can therefore be written as a LRMN, that is, a recurrence relation

$$\begin{aligned}
\hat{\Phi}_t &= \Lambda_t \hat{\Phi}_{t-1} + \Omega_t, && (98)\\
\Lambda_t &= I + g A(\theta) U_{t-1} U_t - g U_{t-1} U_{t-1},\ g \in (0,1), && (99)\\
\Omega_t &= g B(\theta) U_{t-1} U_t, && (100)\\
U_t &= P(\theta) U_{t-1} + \varepsilon_t,\ \varepsilon_t \sim iid(0, \Sigma). && (101)
\end{aligned}$$

Our next step is to argue how a tail index for $\hat{\Phi}_t$ arises in theory. That is, we need to show that the tail of the stationary distribution of $\hat{\Phi}_t$ follows a power law. For this purpose we could appeal to Kesten (1973) if $\Lambda_t$ were $iid$ however only $\varepsilon_t$ is $iid$ so we can write instead

$$\begin{aligned}
\hat{\Phi}_t &= \Lambda_t \hat{\Phi}_{t-1} + \Omega_t, && (102)\\
\Lambda_t &= I + g A(\theta) P(L)^2 L \varepsilon_t \varepsilon_t - g P(L)^2 L^2 \varepsilon_t \varepsilon_t,\ g \in (0,1), && (103)\\
\Omega_t &= g B(\theta) P(L)^2 L^2 \varepsilon_t \varepsilon_t, && (104)\\
U_t &= P(\theta) U_{t-1} + \varepsilon_t,\ \varepsilon_t \sim iid(0, \Sigma) \rightarrow U_t = P(L)\varepsilon_t. && (105)
\end{aligned}$$

While a formal proof that the tail of the stationary distribution of $\hat{\Phi}_t$ and therefore $X_t$ (since it equals $\hat{\Phi}_t U_t$) follows a power law is beyond the scope of this paper (as the mathematics for multivariate correlated processes $U_t$ is not available), we can easily show via simulations

how small realizations of $\varepsilon_t$ can accumulate into small tail indices for model variables. That is, we can observe large fluctuations from trend for model variables even though the model is subject to small shocks.

## 7.5. Estimation Procedure

In this Appendix we detail the method used in our minimum distance estimation of a vector of deep parameters ($\mu$). 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). For example, suppose we wish to only match, by choice of $\mu$, the tail index of HP filtered output ($\approx 4$ 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 $\approx 3$ from Table 1.) then the empirical target is the vector $\varkappa = [4, 3]'$.

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

$$b_t = b_{t-1} + gX_{t-1}(X_t - X'_{t-1}b_{t-1}),\ b_0 \text{ given}, \tag{106}$$
$$X_t = A(\theta)b_{t-1}^2 X_{t-1} + B(\theta)U_t, \tag{107}$$
$$Y_t = H'X_t. \tag{108}$$

   where $X_t$ denotes model variables, $Y_t$ denotes HP filtered data and the deep parameters $\mu$ consist of the constant gain $g$ along with the elements that appear in the matrices $A(\theta)$ and $B(\theta)$; $H'$ 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 $\mu$ 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. Note that the smoother is employed for a given $\mu$.

2. Once the smoothed values $\widehat{X}_t$ are in hand for a given parameterization of $\mu$ 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. Next, suppose that the empirical target is the tail index for HP filtered output.

3. Given a parametrization of $\mu$ 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 $\mu$; that is, our estimates are delivered by

$$\min_{\mu} F = [\varkappa - \varkappa(\mu)]'[\varkappa - \varkappa(\mu)] \tag{109}$$

4. Standard errors are 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, 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.

## 7.6. Standard Error Calculation

Our objective function for estimation of parameters is given by

$$\min_{\mu} F = [\varkappa - \varkappa(\mu)]'[\varkappa - \varkappa(\mu)] \tag{110}$$

where, depending on the estimation exercise in question, we vary $F$ to include more than just the tail indices $\varkappa(\mu)$. Once estimates $(\widehat{\mu})$ are found by minimizing the above function we calculate standard errors by first computing the numerical Hessian $H$ around those estimates. Our standard error vector is then given by the square root of the diagnol elements of $H^{-1}$. We note that in the construction of these standard errors we "balance" the Hessian by constructing a scale vector $(\tau)$ for estimates and then pre and post multiply $H^{-1}$ by the inverse of a matrix consisting of those scale vectors (a standard practice to obtain standard errors when the ranges of estimated parameters vary widely). Moreover, standard errors will be small when estimates are close to starting values (and starting values are themselves converged estimates from another procedure).