

Effective Calibration ¹

A. Ronald Gallant
Department of Economics
University of North Carolina
Chapel Hill NC 27599-3305 USA

July 2001
Last revised: February 2002
Corrections: April 2002

¹Supported by National Science Foundation Grant No. SES0000176. I would like to thank Evan Anderson, Andrew Nobel, and Pranab K. Sen for helpful discussion and comments. The latest version of the paper is available by anonymous ftp at [ftp.econ.duke.edu](ftp://ftp.econ.duke.edu) in directory `pub/arg/papers` as file `edf.ps`. EDF source code and a user's guide are available at the same site in directory `pub/arg/edf`. © 2001 by A. Ronald Gallant.

Abstract

The focus of the paper is the optimal determination from data of the parameters of a structural model that has been rejected by statistical criteria. The structural model is presumed to be amenable to simulation. Rationale that might justify various calibration procedures are discussed and the deficiencies of some standard methods are examined. A calibration procedure based on matching distributions that accords well with the rationale is proposed. Asymptotics are derived under weak assumptions that admit of both time series and cross sectional data without assuming that the process generating data is an instance of the structural model under consideration. The method is illustrated by calibrating an endowment economy to stock returns and consumption data.

JEL Classification: C10, C15, C31, C32

Keywords: Calibration, computational experiments, simulation, empirical processes, alpha-mixing, beta-mixing, edf-calibration.

1 Introduction

This paper relies on a technological advance: Model simulation is now fast enough to make its use in inference practicable. And on a change in perspective: The model is accepted as likely wrong by statistical criteria. The paper then asks the question: Knowing this, what is the best way to determine model parameters from data?

It turns out that much of statistical science, econometrics included, is of little help. The discipline is focused on determining model parameters conditional on the model being correct or detecting the falsity of a model if it is not. At first glance the existence of nonparametric methods would seem to contradict this claim. But nonparametric methods cannot convincingly extrapolate beyond the confines of the data. The wrong model probably can do this. That is usually the reason for interest in it. What it cannot do is fit data in-sample well enough to pass a statistical specification test.

Because the focus of statistical science is elsewhere, its lynchpins – maximum likelihood and method of moments – are not of much help. Moreover, they can be unreliable in the present context because they rely on assumptions that are not necessarily true.

Likelihood based methods are heavily predicated on model correctness and are difficult to justify when that underpinning is removed: Kullback-Liebler discrepancy hardly leaps to mind as the obvious criterion for measuring concordance with the data when the fitted model is wrong. It is not a standard norm with a well developed approximation theory associated with it as has, for instance, the L_2 norm. Likelihood based methods are also apt to be sensitive to observations at the edge of the data cloud. This sensitivity can lead to calibrations that fluctuate markedly from sample to sample or as observations are deleted or added to the ends of a time series.

Method of moments based methods do have intuitive appeal but suffer from three criticisms: (1) the moments that they rely on could easily not exist, (2) calibrations are not invariant to change of scale of measurement in the data, (3) calibrations are not objective because they are based on a researcher's individual preferences.

This criticism of method of moments would collapse immediately if one matched to a collection of bounded moments that had an objective rationale in a way that was invariant

to change in scale of measurement. This observation motivates our approach.

The step functions $r_u(y) = I_{(-\infty, u]}(y)$, where $y \in \mathfrak{R}^d$ and $(-\infty, u]$ is a rectangle in \mathfrak{R}^d , are a basis for all bounded continuous functions of the data. Therefore, a natural calibration procedure would be to choose model parameters to match the moments of step functions for some values of the index u . This is the same as matching the empirical distribution of the data $\hat{F}_n(u)$ to the distribution implied by the model $F(u|\rho)$, where the latter may be computed by simulation. Matching to the entire collection $\{r_u\}_{u \in \mathfrak{R}^d}$ simultaneously can be accomplished by minimizing $\|\hat{F}_n(u) - F(u|\rho)\|$ with respect to some norm $\|\cdot\|$ or other distance function that is analogous to one of the standard norms. The sup norm is not a good choice because simulation methods often introduce granularities that destroy its effectiveness. Some minimal smoothing is necessary. The two natural candidates that provide some smoothing are the L_1 and L_2 norms. Of them, the L_2 is analytically more tractable and is also more traditional. It must be a weighted L_2 norm because distribution functions are not L_2 -Lebesgue. The natural weighting function is the true density of the data. Because the true density of the data is unknown, one integrates with respect to the empirical distribution of the data instead. Integrating with respect to the empirical distribution is equivalent to averaging over the sample $\{y_t\}_{t=1}^n$. These considerations suggest that the appropriate calibration $\hat{\rho}_n$ is the minimizer of the sample objective function

$$s_n(\rho) = \frac{1}{n} \sum_{t=1}^n [\hat{F}_n(y_t) - F(y_t|\rho)]^2.$$

That is our proposal. We call it edf-calibration.

Another rationale that would lead to edf-calibration is the following: Where the model makes contact with the data, that is, over the support of the data, it is natural to try to make probabilities computed with respect to the model as close to those implied by the data as possible. Since errors are inevitable, it is natural to penalize errors where the data are abundant the most. This rationale leads to the same proposal.

As one would expect, and as is shown in Section 3 under weak regularity conditions that accommodate both time series and cross sectional data, $\hat{\rho}_n$ tends to the parameter value ρ^o that minimizes

$$s^o(\rho) = \int [F(y) - F(y|\rho)]^2 f(y) dy,$$

where $F(y)$ and $f(y)$ are the distribution and density functions, respectively, that actually govern the data. This characterizes the population counterpart ρ^o of edf-calibration and shows that the rationale advanced for the sample calibration carries over to the population calibration. Section 3 also shows that $\sqrt{n}(\hat{\rho}_n - \rho^o)$ is asymptotically normal under stronger regularity conditions. These are the main theoretical contributions of the paper.

In the remaining sections, edf-calibration is defined, asymptotics are derived, minimization algorithms are discussed, and examples are presented. A final section concludes. The rest of this section reviews the literature related to the proposed calibration procedure and presents arguments supporting the claims above.

Calibration means different things to different people as the exchange among Kydland and Prescott (1996), Hansen and Heckman (1996), and Sims (1996) illustrates quite clearly. This debate ranges over what calibration is, whether it ought to be done, and how it should be done. Geweke (1999) augments this exchange with a more technical elaboration of the ideas and illustrates with interesting examples. This same debate occurs in other disciplines. For instance, the debate between Caswell (1988) and Hall (1988) that occurred in *Ecological Modelling* nearly reproduces the debate above except for the substitution of ecological models for economic models and less concern for the precepts of statistical science. In view of the differing perspectives evidenced in these writings, it is best to be precise by what is meant by calibration in this paper.

What is meant here is that one needs data to determine some model parameters and one is deliberately fitting a rejected statistical model to that data to determine them. One reason for doing this that is advanced in the literature is the following: One must extrapolate well beyond available data for some purpose. One argues that reduced form models that produce good in-sample fits can be unreliable at large distances from the data cloud. Reliance on an internally consistent structural model is quite likely to be safer, whether it fits in-sample or not. Another reason that one might deliberately fit a rejected model to data is to determine model parameters in a reasonably objective way so as to permit a study of the properties of the model. Whatever the reason, calibration from data is frequently done and it is of interest to consider the merits of various methods for doing so.

The context needs to be established. We are primarily concerned here with structural

models that can be simulated and for which methods that rely on simulation are appropriate. This is usually because analytical intractability precludes the use of other methods. The presence of unobserved state variables is the most common cause of this intractability which, for instance, precludes the use of generalized method of moments. As the example of Section 5 illustrates, it may be necessary to introduce unobserved states for the purpose of calibration even for models that might not otherwise require simulation estimators.

If the structural model actually does fit the data, there are good alternatives to our proposal within the context. One can use simulated maximum likelihood; see Durham and Gallant (2001) and the references therein. One can use methods that are asymptotically equivalent to quasi maximum likelihood such as efficient method of moments and indirect inference; see Gourieroux, Monfort and Renault (1993) or Gallant and Tauchen (2001) and the references therein. One can use simulation based Bayesian methods; see Eraker (2001) and the references therein. One can use simulated method of moments, see Duffie and Singleton (1993) and the references therein. Outside of the context, other methods become available, but we are not concerned with them.

To elaborate on the problems presented by likelihood based calibration, suppose that the structural model determines a density $f(y|\rho)$ for iid data $\{y_t\}$ that actually follows a distribution with density $f(y)$. If a simulation based equivalent of quasi maximum likelihood is used to determine model parameters then the calibration converges to the minimizer $\rho^\#$ of the Kullback-Liebler discrepancy

$$\int [\log f(y) - \log f(y|\rho)] f(y) dy.$$

If $f(y)$ is not in the family of densities indexed by the parameter ρ , then it is not at all clear that minimizing Kullback-Liebler discrepancy is a sensible way to calibrate model parameters. No useful approximation properties flow from this measure of distance under these circumstances. Under certain regularity conditions, relations that imply continuity of standard norms with respect to Kullback-Liebler discrepancy such as

$$\int |f(y) - f(y|\rho)| dy \leq C \left[\int [\log f(y) - \log f(y|\rho)] f(y) dy \right]^{1/2}$$

can be established. While inequalities of this sort are useful in establishing convergence rates for density estimates when Kullback-Liebler discrepancy tends to zero, they are of little value

in characterizing the properties of a calibration when Kullback-Liebler discrepancy tends to a positive constant as it does here.

Kullback-Liebler discrepancy places heavy emphasis on approximating the tails of the density well. In a calibration situation, it would seem that one would be more interested in approximating well in the center of the distribution, or at least in trying to be somewhat even handed. This attention to tails can have deleterious consequences. Structural models tend to be better behaved than actual data so that quite often $f(y|\rho)$ will have thinner tails than $f(y)$. For an example, see Figure 3. As is well known (Huber, 1981), when this happens one can expect quasi maximum likelihood calibrations to be sensitive to observations at the edges of the data cloud and therefore to be unstable, varying markedly from sample to sample and exhibiting large fluctuations when the time span of a sample is changed. It would seem from these observations that quasi maximum likelihood and its asymptotic equivalents should be avoided for the purpose of calibration.

To elaborate on the problems presented by simulated method of moments calibration, let the densities $f(y)$ and $f(y|\rho)$ be defined as above. A method of moments calibration converges to the minimizer $\rho^\#$ of

$$\left\| \int \psi(y)f(y) dy - \int \psi(y)f(y|\rho) dy \right\|$$

where $\| \cdot \|$ is some measure of distance and ψ some vector of moment functions, both selected subjectively by the calibrator. Subjectivity can be a vice or a virtue; we shall not argue that point at length. This author thinks it to be a vice. The integral $\int |\psi(y)| f(y) dy$ on which the method relies may not be finite. When that happens, the method fails and calibrations can vary markedly from sample to sample and exhibit large fluctuations when the time span of a sample is changed. The application of Section 5 is an instance where one might be concerned about the reliability of method of moments calibrations because the empirical results of Müller, Dacorogna, and Picter (1998) and citations therein indicate that a presumption of existence of moments past the second is dubious for data from financial markets. But why take a chance on the existence of any moment when the assumption is unnecessary?

A structural problem with method of moments calibration is a lack of invariance to

change in scale of measurement. For example, suppose the structural model implies log gross returns are normally distributed with location parameter μ and scale parameter σ whereas the observed log returns data actually follow the Laplace distribution with location parameter ν and scale parameter τ , which is actually a plausible assumption because the Laplace has thicker tails and is more peaked than the normal as are most returns data. Suppose one matches first and second raw moments. If the match is done in log returns then, as sample size increases, the calibration converges to

$$\begin{aligned}\mu &= \nu \\ \sigma^2 &= 2\tau^2.\end{aligned}$$

If the match is done in returns and $\tau < \frac{1}{2}$ then the calibration converges to

$$\begin{aligned}\mu &= \nu - 2\log(1 - \tau^2) + \frac{1}{2}\log(1 - 4\tau^2) \\ \sigma^2 &= 2\log(1 - \tau^2) - \log(1 - 4\tau^2).\end{aligned}$$

If the match is done in returns and $\tau \geq \frac{1}{2}$ then the second moment does not exist and the calibration fails.

To some extent, it is difficult to contrast with Bayesian methods. As Sims (1996) points out, the Bayesian development is axiomatic. If one accept the axioms, then one must accept the method and there is nothing to discuss. Amusingly, as Sims also points out, much of science muddles through without regard for the Bayesian axioms, or for the rest of statistical science for that matter. One can argue, and Sims makes this point in a sense, that the calibration activity as described here is outside the scope of formal statistical inference and is more in the nature of a numerical computation for the purpose of assessing the features of an economic model. If this view is adopted, one is outside the scope of the axioms that justify the Bayesian system of inference.

In the context established here the practical problem posed by a Bayesian approach is that one must actually envisage an exhaustive collection of models of which the structural model under consideration is a member. Then one can make comparative probability statements and give probabilistic estimates of the reliability of a estimate. Geweke suggests a reduced form model to augment the class of models so that the requisite encompassing is plausible.

But probability statements relative to a reduced form model seem hardly in the spirit of what we hope to accomplish here. Without reduced form augmentation, one must envisage all possible candidate structural models that one would be willing to entertain, which is an impossible task. The only alternative to a collection of models that exhaust all possibilities within the Bayesian paradigm is acceptance of the structural model under consideration as correct for the purpose of calibrating it. One is now conditioning inference on a model that is effectively known to be false. This seems internally inconsistent.

Philosophical issues aside, if one just views the Bayesian methodology as a possible calibration device, problems still remain. Aside from writing the simulator, which must be done regardless, implementation of edf-calibration is a straightforward optimization problem that can be implemented once and for all. In contrast, the Bayesian method requires considerable extra effort that must be tailored to each application to devise a (MCMC) integration strategy to eliminate unobserved states and to obtain the posterior. One cannot achieve objectivity with Bayesian methods. There is no truly flat or uninformative prior. Any prior thought to be uninformative can be revealed to be informative by a one-to-one transformation of the parameters of the model. Moreover, if one actually does make a serious attempt to give every parameter value equal chance under the prior, then the stability problems of quasi maximum likelihood mentioned above become an issue for the same reasons.

In the view of Kydland and Prescott (1996), matching to the the entire distribution $F(y)$ is too stringent a requirement. One's intention is only to match some population notion of central tendency and nothing more. In this case, as Geweke (1999) argues, a model must be further augmented to provide a point of contact between the measure of central tendency and the distribution of the data. Geweke's argument that there is no logical escape from this step is persuasive. An exploration along these lines is in Section 5 where a process is implicitly interposed between the endowment process and the consumption process.

Inference strategies based on matching the empirical distribution function to the distribution function implied by a model are not new. The Kolmogorov-Smirnov lack of fit test is the obvious example. The objective function $s_n(\rho)$ proposed here is similar to the Cramer-von Mises test statistic; the difference between them is that $s_n(\rho)$ weights by the empirical distribution of the data whereas the Cramer-von Mises test statistic weights by the

hypothesized null distribution. For univariate data, an estimator obtained by minimizing the Cramer-von Mises test statistic, was proposed by Parr and De Wet (1981), with the null taken to be $F(y|\rho)$ in the notation of this paper. Parr and De Wet emphasize its robustness properties, show how to custom tailor its influence function, and argue that, because the fit is in units of probability, it is better suited to fitting a misspecified model to data than other estimation strategies such as M- and L-estimators. Variants of the Cramer-von Mises statistics have been proposed in the statistics literature that weight by null or non-null empirical distribution functions. These are more closely related to $s_n(\rho)$ as proposed here; to our knowledge they have not been used to define an estimator. In the econometrics literature, Doraszelski and Matzkin (2001) and Matzkin (2001) contain estimation strategies based on variants of $s_n(\rho)$ above; they, in turn, cite Manski (1983).

What is new here is the proposal that $s_n(\rho)$ is an effective calibration tool for multi-dimensional models with latent variables when implemented by means of simulation, the development of computational algorithms, and the derivation of multidimensional asymptotics with misspecification presumed under weak mixing conditions.

2 EDF-Calibration

The components of edf-calibration are the empirical distribution function of the data $\hat{F}_n(y)$, the distribution function implied by the model $F(y|\rho)$, its numerical approximant $\hat{F}_N(y|\rho)$, the sample objective function $s_n(\rho)$, and its minimizer $\hat{\rho}_n$, which is the calibration. These are defined in turn below.

DEFINITION 1 Write $I(u)$ for the univariate distribution function that puts all its mass at zero; that is, $I(u) = I_{[0,\infty)}(u)$ for $u \in \mathfrak{R}^1$. In terms of $I(\cdot)$, the empirical distribution function of a sequence of vectors in \mathfrak{R}^d

$$y_t = (y_{1,t}, \dots, y_{d,t}) \quad t = 1, \dots, n$$

is

$$\hat{F}_n(v) = \frac{1}{n} \sum_{t=1}^n \prod_{i=1}^d I(v_i - y_{it})$$

for $v \in \mathfrak{R}^d$.

DEFINITION 2 Let $F(v|\rho)$ be the distribution function implied by the structural model.

It can be approximated by

$$F_N(v|\rho) = \frac{1}{N} \sum_{t=1}^N \prod_{i=1}^d I(v_i - \hat{y}_{it})$$

where

$$\hat{y}_t = (\hat{y}_{1,t}, \dots, \hat{y}_{d,t}) \quad t = 1, \dots, N$$

is a simulation from the structural model with parameters set to ρ .

DEFINITION 3 For a sample

$$y_t = (y_{1,t}, \dots, y_{d,t}) \quad t = 1, \dots, n$$

define the sample objective function by

$$s_n(\rho) = \frac{1}{n} \sum_{t=1}^n [\hat{F}_n(y_t) - F(y_t|\rho)]^2.$$

DEFINITION 4 Define $\hat{\rho}_n$ to be an edf-calibration of the structural model if

$$\hat{\rho}_n = \underset{\rho \in R}{\operatorname{argmin}} s_n(\rho)$$

where $R \subset \mathfrak{R}^p$ is the set of admissible parameter values of the structural model.

Edf-calibration is invariant to a change in scale of measurement as shown next.

LEMMA 1 If the distribution function implied by the structural model is continuous, then edf-calibration is invariant to invertible, increasing, differentiable transformations that are applied elementwise.

Proof Let ϕ denote the map $y \mapsto [\phi_1(y_1), \dots, \phi_d(y_d)]$, where each ϕ_i is invertible, differentiable, and increasing. In terms of the density of the structural model for the original data $f_Y(y|\rho)$, the transformed data $v_t = \phi(y_t)$ has distribution

$$\begin{aligned} F_V(u|\rho) &= \int \left(\prod_{i=1}^d I(u_i - v_i) \right) f_Y(\phi^{-1}(v) | \rho) \left| \det \left(\frac{\partial}{\partial v'} \phi^{-1}(v) \right) \right| dv \\ &= \int \prod_{i=1}^d I[u_i - \phi_i(y_i)] f_Y(y|\rho) dy \\ &= \int \prod_{i=1}^d I[\phi_i^{-1}(u_i) - y_i] f_Y(y|\rho) dy \\ &= F_Y[\phi^{-1}(u) | \rho]. \end{aligned}$$

In terms of the empirical distribution \hat{F}_{Y_n} of the original data, the empirical distribution of the transformed data is

$$\begin{aligned}
\hat{F}_{V,n}(u) &= \frac{1}{n} \sum_{t=1}^n \prod_{i=1}^d I(u_i - v_{it}) \\
&= \frac{1}{n} \sum_{t=1}^n \prod_{i=1}^d I[u_i - \phi_i(y_{it})] \\
&= \frac{1}{n} \sum_{t=1}^n \prod_{i=1}^d I[\phi_i^{-1}(u_{it}) - y_{it}] \\
&= \hat{F}_{Y,n}[\phi^{-1}(u)].
\end{aligned}$$

Then,

$$\begin{aligned}
s_{V,n}(\rho) &= \frac{1}{n} \sum_{t=1}^n [\hat{F}_{V,n}(v_t) - F_V(v_t|\rho)]^2 \\
&= \frac{1}{n} \sum_{t=1}^n \left\{ \hat{F}_{Y,n}[\phi^{-1}(v_t)] - F_Y[\phi^{-1}(v_t)|\rho] \right\}^2 \\
&= \frac{1}{n} \sum_{t=1}^n [\hat{F}_{Y,n}(y_{it}) - F_Y(y_{it}|\rho)]^2 \\
&= s_{Y,n}(\rho).
\end{aligned}$$

□

3 Asymptotics

The use of mixing to express limits on dependence is standard in the dynamic nonlinear models literature, e.g. Gallant and White (1987), Davidson (1994), Pötscher and Prucha (1996). Of the extant mixing concepts, α -mixing is the weakest (Doukhan, 1994). Initially we shall use α -mixing to express limits on dependence but later, in order to have access to results that rely on symmetrization arguments, we shall have to impose the stronger concept of β -mixing. We remark that iid data satisfies both the α - and β -mixing assumptions below so that the results here apply to both cross sectional and time series data.

ASSUMPTION 1 Let $\{y_t\}_{t=-\infty}^{\infty}$ be a sequence of random variables defined on a complete probability space (Ω, \mathcal{A}, P) with range in \mathfrak{R}^d that is α -mixing of size $-r/(r-2)$ for some $r > 2$. Let $\{y_t\}$ be stationary with continuous distribution function $F(y)$ that has density $f(y)$.

To elaborate, a measure of dependence between two σ -algebras \mathcal{F} and \mathcal{G} is

$$\alpha(\mathcal{F}, \mathcal{G}) = \sup_{F \in \mathcal{F}, G \in \mathcal{G}} |P(FG) - P(F)P(G)|.$$

The measure is zero if the two σ -algebras are independent and is positive otherwise. The σ -algebra containing the information on the past is $\mathcal{F}_{-\infty}^t = \sigma(y_t, y_{t-1}, \dots)$; the σ -algebra containing the information on the future after a gap m is $\mathcal{F}_{t+m}^\infty = \sigma(y_{t+m}, y_{t+m+1}, \dots)$. Let

$$\alpha_m = \sup_t \alpha(\mathcal{F}_{-\infty}^t, \mathcal{F}_{t+m}^\infty).$$

The rate at which α_m converges to zero as the gap m increases is a measure of the dependence in the sequence $\{y_t\}$; faster rates imply less dependence. In the mixing literature, this rate is often expressed in terms of size: $\{y_t\}$ is said to be α -mixing of size $-q$ if $\alpha_m = O(m^{-q-\delta})$ for some $\delta > 0$. Doukhan (1994) and Withers (1981) contain examples of such processes; α -mixing is a plausible assumption for the output of a dynamical system.

The stationarity and dependence conditions imposed by Assumption 1 are sufficient to establish the Glivenko-Cantelli result:

LEMMA 2 Assumption 1 implies

$$\lim_{n \rightarrow \infty} \sup_y |\hat{F}_n(y) - F(y)| = 0$$

almost surely.

Proof We apply Theorem 1 of Gallant (1987a, Chapter 7) with $V_t = y_t$, with W_t the projection map $V_\infty \mapsto y_t$ and with $g_t(W_t, \gamma) = \prod_{i=1}^d I(v_i - y_i)$ for given $v \in \mathfrak{R}^d$. Because W_t is a projection map and the dependence of g_t on t and γ is trivial, g_t is near epoch dependent and A-smooth. The dominating sequence is $d_t(\omega) \equiv 1$. We conclude that $\hat{F}_n(v) \rightarrow F(v)$ a.s. for each fixed v . Equivalently, $(1/n) \sum_{t=1}^n r(y_t) \rightarrow \int r(u) dF(u)$ a.s. for each r in the collection \mathcal{R} of indicator functions of rectangles in \mathfrak{R}^d of the form $(-\infty, v] = \prod_{i=1}^d (-\infty, v_i]$. Nobel and Olshen (1996, Appendix III) construct an ϵ -bracketing for the collection of indicator functions of half planes in \mathfrak{R}^d that makes use of the assumption that F has a density. That proof applies equally as well to \mathcal{R} because one only needs to substitute the boundary of a rectangle for the boundary of a half plane in their argument. The only modification required is that

their bound on the number of small cubes required to cover the intersection of a sphere and boundary of a half plane must be increased by the factor d to get a bound on the number of small cubes required to cover the intersection of a sphere and boundary of a rectangle. Apply Theorem 2 of Pollard (1984, p. 8) to conclude that $\sup_{r \in \mathcal{R}} |(1/n) \sum_{t=1}^n r(y_t) - \int r(u) dF(u)| \rightarrow 0$ a.s.; equivalently, to conclude that $\lim_{n \rightarrow \infty} \sup_v |\hat{F}_n(v) - F(v)| = 0$ a.s. \square

The structural model will be required to satisfy the same stationarity and dependence conditions that were imposed on the true data generating process in Assumption 1; in addition, a smoothness condition is imposed:

ASSUMPTION 2 For each ρ in some compact parameter space R , which is a subset of \mathfrak{R}^p , the data generating process implied by the structural model is α -mixing of size $-r/(r-2)$ for some $r > 2$ and is stationary with continuous distribution function $F(y|\rho)$ that has density $f(y|\rho)$. Moreover, there is a dominating function $D(y)$ with $D^r(y)$ integrable with respect to the distribution $F(y)$ given by Assumption 1 such that $|(\partial/\partial\rho_i)F(y|\rho)| \leq D(y)$ for $i = 1, \dots, p$.

The compactness assumption in Assumption 2 is standard in the nonlinear models literature. It can be eliminated on an ad hoc basis by exploiting the features of a particular application to show that the minimizer $\hat{\rho}_n$ must be contained in a compact set for large n . Most practitioners seem to take the attitude that one must effectively know R in order to use minimization routines so that one might as well make use of this same knowledge in the theoretical development.

In a sense that the results in the remainder of this section makes precise, edf-calibration has a population counterpart that is expressed in terms of the following objects. The terminology in the next definition anticipates these results.

DEFINITION 5 The population objective function is

$$s^o(\rho) = \int [F(y) - F(y|\rho)]^2 f(y) dy.$$

Its minimum

$$\rho^o = \operatorname{argmin}_{\rho \in R} s^o(\rho)$$

is the population calibration.

In large samples, the population and sample objective functions are indistinguishable:

LEMMA 3 Assumptions 1 and 2 imply

$$\lim_{n \rightarrow \infty} \sup_{\rho \in R} |s_n(\rho) - s^o(\rho)| = 0$$

almost surely; $s^o(\rho)$ is continuous on R .

Proof

$$\begin{aligned} & \sup_{\rho \in R} \left| s_n(\rho) - \frac{1}{n} \sum_{t=1}^n [F(y_t) - F(y_t|\rho)]^2 \right| \\ &= \frac{1}{n} \sum_{t=1}^n [\hat{F}_n(y_t) - F(y_t)]^2 + \sup_{\rho \in R} \left| \frac{2}{n} \sum_{t=1}^n [\hat{F}_n(y_t) - F(y_t)] [F(y_t) - F(y_t|\rho)] \right| \\ &\leq \sup_y [\hat{F}_n(y) - F(y)]^2 + 4 \sup_y |\hat{F}_n(y) - F(y)| \end{aligned}$$

The two terms following the inequality converge to zero a.s. by Lemma 2. We now apply Theorem 1 of Gallant (1987a, Chapter 7) to $(1/n) \sum_{t=1}^n [F(y_t) - F(y_t|\rho)]^2$ with $V_t = y_t$, with W_t the map $V_\infty \mapsto y_t$, with $\gamma = \rho$, and with $g_t(W_t, \gamma) = F(y) - F(y|\rho)$. Because W_t is a projection map, g_t is trivially near epoch dependent. The A-smooth condition is implied by Taylor's theorem and the domination of the partial derivatives $(\partial/\partial\rho_i)F(y|\rho)$ by $D(y)$. The dominating sequence for $g_t(W_t, \gamma)$ is $d_t(\omega) \equiv 2$. Upon invoking Theorem 1,

$$\lim_{n \rightarrow \infty} \sup_{\rho \in R} \left| \frac{1}{n} \sum_{t=1}^n [F(y_t) - F(y_t|\rho)]^2 - \int [F(v) - F(v|\rho)]^2 f(v) dv \right|$$

and the conclusion of the lemma follows. □

We impose an identification condition:

ASSUMPTION 3 The minimum ρ^o of $s^o(\rho)$ over R is unique.

Assumption 3 can be relaxed provided that convergence results are also modified along the lines of the statistical literature for linear models of less than full rank. See Gallant (1987b) for details. The basic point is that any calibration $\rho^\#$ that produces the same distribution as $F(y|\rho^o)$ can be accepted as being useful for many purposes because certain downstream computations are invariant. At the present juncture, trying to establish results at this level of generality does not seem to add anything of practical value to our understanding. Our result is, then, that in large samples, the sample and population calibrations are the same:

THEOREM 1 Assumptions 1 through 3 imply

$$\lim_{n \rightarrow \infty} \hat{\rho}_n = \rho^o$$

almost surely .

Proof The argmin function is continuous with respect to the uniform convergence of a function $s_n(\rho)$ to a continuous function $s^o(\rho)$ on a compact set when the minimum of $s^o(\rho)$ is unique. Assumption 3 and the conclusion of Lemma 3 supply these conditions. \square

We remark that a consequence of Theorem 1 and Lemma 1 taken together is that the population calibration ρ^o is invariant to changes in scale of measurement.

To determine the sampling variability of an edf-calibration, it is necessary to replace Assumption 1 with the following more restrictive assumption.

ASSUMPTION 4 Let $\{y_t\}_{t=-\infty}^{\infty}$ be a sequence of random variables defined on a complete probability space (Ω, \mathcal{A}, P) with range in \mathfrak{R}^d that is β -mixing of size $-r/(r-2)$ for some $r > 2$. Let $\{y_t\}$ be stationary with $y_t \sim F(y)$ and $(y_s, y_t) \sim F_{s,t}(u, v)$, where $F(y)$ has density $f(y)$. Let $F(y|\rho)$ be as given by Assumption 2 and define

$$k(u, v) = \left[\prod_{i=1}^d I(v_i - u_i) - F(v|\rho^o) \right] \frac{\partial}{\partial \rho} F(v|\rho^o).$$

Let $\int |\lambda' k(u, v)|^r dF_{s,t}(u, v)$, $\int |\lambda' k(v, u)|^r dF_{s,t}(u, v)$, and $\iint |\lambda' k(u, v)|^r dF(u) dF(v)$ be finite for all $s < t$ and all $\lambda \in \mathfrak{R}^d$ with $\lambda' \lambda = 1$. Assume that the partial derivatives of $F(y|\rho)$ with respect to ρ up to the third order are dominated by a function $D(y)$ for which $D^r(y)$ is integrable with respect to $F(y)$; e.g. $\sup_{\rho \in R} |(\partial^2 / \partial \rho_i \partial \rho_j) F(y|\rho)| \leq D(y)$, where R is as in Assumption 2. Assume ρ^o is an interior point of R .

To elaborate, the β -mixing coefficient is

$$\beta_m = \sup_t \mathcal{E} \left\{ \sup_{F \in \mathcal{F}_{t+m}^{\infty}} \left| P(F|\mathcal{F}_{-\infty}^t) - P(F) \right| \right\},$$

where $\mathcal{F}_{-\infty}^t$ is the σ -algebra containing the information on the past and $\mathcal{F}_{t+m}^{\infty}$ is the σ -algebra containing the information on the future after a gap m , as defined earlier. β -mixing of given size implies α -mixing of the same size because $\alpha_m \leq \beta_m$ (Doukhan, 1994).

THEOREM 2 Assumptions 2 through 4 imply that $\sqrt{n}(\hat{\rho}_n - \rho^\circ)$ is asymptotically normally distributed.

Proof The first order conditions for the sample objective function are

$$0 = \frac{\partial}{\partial \rho} s_n(\hat{\rho}_n) = \frac{-2}{n} \sum_{t=1}^n [\hat{F}_n(y_t) - F(y_t|\hat{\rho}_n)] \frac{\partial}{\partial \rho} F(y_t|\hat{\rho}_n)$$

and for the population objective function they are

$$0 = \frac{\partial}{\partial \rho} s^\circ(\rho^\circ) = -2 \int [F(y) - F(y|\rho^\circ)] \frac{\partial}{\partial \rho} F(y|\rho^\circ) dF(y),$$

where the domination condition on partial derivatives permits the interchange of integration and differentiation. A Taylor's expansion of the sample first order conditions gives

$$\begin{aligned} & \mathcal{J}_n(\bar{\rho}_n) \sqrt{n}(\hat{\rho}_n - \rho^\circ) \\ &= \frac{-2}{\sqrt{n}} \sum_{t=1}^n [\hat{F}_n(y_t) - F(y_t|\rho^\circ)] \frac{\partial}{\partial \rho} F(y_t|\rho^\circ) \\ &= \frac{-2}{\sqrt{n}} \sum_{t=1}^n \left[\frac{1}{n} \sum_{s=1}^n \prod_{i=1}^d I(y_{it} - y_{is}) - F(y_t|\rho^\circ) \right] \frac{\partial}{\partial \rho} F(y_t|\rho^\circ) \\ &= \frac{-2}{n\sqrt{n}} \sum_{t=1}^n \sum_{s=1}^n \left[\prod_{i=1}^d I(y_{it} - y_{is}) - F(y_t|\rho^\circ) \right] \frac{\partial}{\partial \rho} F(y_t|\rho^\circ) \\ &= \frac{1}{n\sqrt{n}} \frac{\partial}{\partial \rho} \sum_{t=1}^n \sum_{s=1}^n \left[\prod_{i=1}^d I(y_{it} - y_{is}) - F(y_t|\rho^\circ) \right]^2 \\ &= \frac{1}{n\sqrt{n}} \frac{\partial}{\partial \rho} \sum_{t=1}^n [1 - F(y_t|\rho^\circ)]^2 + \frac{1}{n\sqrt{n}} \frac{\partial}{\partial \rho} \sum_{s \neq t}^n \left[\prod_{i=1}^d I(y_{it} - y_{is}) - F(y_t|\rho^\circ) \right]^2 \\ &= \frac{1}{\sqrt{n}} A_n - \frac{2(n-1)}{n} \sqrt{n} U_n \end{aligned}$$

where

$$\begin{aligned} \mathcal{J}_n(\rho) &= \frac{-2}{n} \sum_{t=1}^n \left\{ \frac{\partial}{\partial \rho} F(y_t|\rho) \frac{\partial}{\partial \rho'} F(y_t|\rho) - [\hat{F}_n(y_t) - F(y_t|\rho)] \frac{\partial^2}{\partial \rho \partial \rho'} F(y_t|\rho) \right\} \\ U_n &= \frac{1}{n(n-1)} \sum_{s \neq t}^n \frac{1}{2} [k(y_s, y_t) + k(y_t, y_s)] \end{aligned}$$

and $\bar{\rho}_n$ is a point on the line segment joining $\hat{\rho}_n$ to ρ° . The elements of $\mathcal{J}_n(\rho)$ are actually evaluated at different such $\bar{\rho}_n$, but, because the number of different $\bar{\rho}_n$ is finite, there is no need to keep track of them with a cluttered notation. By Theorem 1 of Gallant (1987a,

Chapter 7), $\mathcal{J}_n(\rho)$ converges a.s. uniformly on R to a nonrandom matrix $\mathcal{J}^o(\rho)$ with continuous elements. By Theorem 1 above, $\hat{\rho}_n$ converges to ρ^o a.s. whence $\bar{\rho}_n$ must also because it is on the line segment joining $\hat{\rho}_n$ to ρ^o . Therefore, $\mathcal{J}_n(\bar{\rho}_n)$ converges a.s. to $\mathcal{J}^o(\rho^o)$. By Slutsky's theorem, it is now only necessary to show that the expression after the last equal sign of the first order conditions is asymptotically normal.

Yoshihara (1976) shows that $\sqrt{n}[\lambda'U_n - \iint \lambda'k(u, v) dF(u)dF(v)]$ is asymptotically normal under Assumption 4. Arcones and Yu (1994) give an empirical process version of the same result with k indexing the empirical process. Now

$$\iint k(u, v) dF(u)dF(v) = \int [F(v) - F(v|\rho^o)] \frac{\partial}{\partial \rho} F(v|\rho^o) dF(v)$$

and the population first order conditions imply that the right hand side of this equality is zero. Therefore, $\sqrt{n}\lambda'U_n$ is asymptotically normal. Because λ is arbitrary, $\sqrt{n}U_n$ is asymptotically normal. Now A_n converges a.s. by Theorem 1 of Gallant (1987a, Chapter 7) so that $(1/\sqrt{n})A_n$ converges a.s. to zero. Therefore, the expression after the last equal sign of the first order conditions is asymptotically normal. \square

We shall not attempt to characterize and estimate the variance of the limiting distribution of the calibration, namely $\text{Var}[\sqrt{n}(\hat{\rho}_n - \rho^o)]$, which is a formidable task, because it cannot be relied on to accurately represent sampling variation for two reasons: (1) the choice of simulation method may affect the accuracy of the asymptotics, and (2) there may be additional variation not reflected in the asymptotics due to numerical instability in computations.

To elaborate, economic systems are difficult to solve and simulate. Even if done very well, there still can be some granularity in simulated y_t caused by discretization. The Markov chain method that we used in Section 5 is an example, although we were careful to increase the number of states until results stabilized. This slight departure from the regularity conditions, which presume continuity, may cause $\text{Var}[\sqrt{n}(\hat{\rho}_n - \rho^o)]$ to poorly approximate variability. Issues of granularity aside, even simulation methods that do not discretize can exhibit numerical instabilities that may contribute to variability. In addition, instability in the minimization algorithm employed may contribute to variability. Determination of calibration variability by bootstrap methods takes all these sources of variability into account and is therefore the preferred approach.

In the work reported here, we use the method proposed by Politis and Romano (1994). Briefly, their proposal is as follows. In order to allow indexes to exceed the sample size n , append a copy of the data to the data. Choose a beginning index i by sampling the uniform distribution on the integers 1 through n . Choose a block length m by sampling the geometric distribution $(1 - q)q^m$, where $m = 0, 1, \dots$. If m exceeds $n - 1$ put m to $n - 1$. The terminating index of the block is $i + m$. The bootstrap sample consists of such blocks laid end to end until a data set of size n is reached. The parameter q controls the block size, where $0 \leq q < 1$. Larger q gives larger block sizes; $q = 0$, which is taken to mean that $m = 0$ with probability one, is appropriate for iid data. In our work, we choose q to maximize the reported estimate of scale. Our preferred measure of scale for an element of ρ is the interquartile range, $\text{IQR} = Q_3 - Q_1$, where Q_1 and Q_3 are the first and third quartiles of its bootstrap distribution.

4 Computations

Minimizing

$$s_n(\rho) \doteq \frac{1}{n} \sum_{t=1}^n \left[\hat{F}_n(y_t) - F_N(y_t|\rho) \right]^2.$$

is not an easy problem because $s_n(\rho)$ is not a smooth function of ρ due to the indicator functions that appear in the definition of $\hat{F}_n(y)$ and $F_N(y|\rho)$.

The problem is harder still if the simulation strategy granularizes the data. The application in Section 5 is an illustration. The structural model there is solved by replacing a continuous transition density with a Markov chain approximation. This discretizes the state space which, in turn, discretizes the simulated distribution $F_N(y|\rho)$. The situation is shown in Figure 1 for equity returns data.

Figure 1 about here

Smoothing the distributions $\hat{F}_n(y)$ and $F_N(y|\rho)$ prior to matching stabilizes the computations considerably. The smoothed problem is fairly easy to solve. The solution to the smoothed problem can either be accepted as the solution to the original problem or used

as a start value for solving the more difficult original problem. An effective smoother is the quadratic squasher defined as follows.

DEFINITION 6 The distribution function

$$S(u) = \frac{u^2 + u|u| + 2u + 2|u| + 4}{2u^2 + 4|u| + 8}$$

closely approximates the logistic distribution but is much cheaper to compute because the exponential function does not need to be evaluated. We term this function a quadratic squasher following the neural net literature. The scaled version is

$$S_\sigma(u) = S\left(\frac{u}{\sigma}\right),$$

which is symmetric in the sense $S_\sigma(-u) = 1 - S_\sigma(u)$ and satisfies

$$\lim_{\sigma \rightarrow 0} S_\sigma(u) = \begin{cases} 0 & u < 0 \\ 1/2 & u = 0 \\ 1 & u > 0 \end{cases}$$

DEFINITION 7 A smooth approximation to the empirical distribution $\hat{F}_n(y)$ is

$$\hat{F}_{\sigma,n}(v) = \frac{1}{n} \sum_{t=1}^n \prod_{i=1}^d S_\sigma(v_i - y_{it}).$$

For a simulation $\{\hat{y}_t\}_{t=1}^N$ from the structural model with parameters set to ρ , a smooth approximation to $F_N(y|\rho)$ is

$$F_{\sigma,N}(v|\rho) = \frac{1}{N} \sum_{t=1}^N \prod_{i=1}^d S_\sigma(v_i - \hat{y}_{it}).$$

Plots of the quadratic squasher for different scaling factors σ are shown in Figure 2. The smoothing effect of the squasher is shown in Figure 3. Figure 3 is Figure 1 after smoothing.

Figure 2 about here

Figure 3 about here

The smoothed problem can be solved by standard nonlinear optimizers. A method for solving the original problem is to evaluate $s_n(\rho)$ over thousands of randomly selected points in a ball that has the smoothed solution as its center, accepting the minimum of these as the answer. The unsmoothed objective function seems to be locally smooth because it pays to let the optimizer polish each of these randomly selected points with a few iterations.

One could view the solution of the smoothed problem as a calibration in its own right and develop an asymptotic theory to accompany it, which would actually be quite straightforward following the same development as in Section 3. But computations for the original problem, while burdensome, are not prohibitively time consuming. One would probably waste more time trying to determine the smoothing parameter σ optimally than it would have taken to optimize $s_n(\rho)$ directly using the algorithm suggested here.

5 Application

We illustrate ideas by calibrating a classical model to classical data: We shall calibrate a Lucas (1978) representative agent endowment economy to the data of Hansen and Singleton (1982, 1983, 1984). For numerical stability, we shall use a variant of the model proposed by Mehra and Prescott (1985). The solution method is due to Tauchen and Hussey (1991).

When the agent possess separable constant relative risk aversion utility, as assumed here, this type of model is known to fit joint returns and consumption growth data poorly. To calibrate it, we shall follow a fairly standard practice of first calibrating the endowment process by calibrating to consumption data alone and then, holding the parameters of the endowment process fixed at these values, calibrate the remaining parameters from equity returns. Log consumption growth is modeled as a first order autoregression. We calibrate three variants: Model 1 accepts the consumption process as the endowment process and calibrates the utility function parameters. Model 2 allows the scale parameter of the endowment process to adjust as well as utility function parameters. Model 3 is used to calibrate a latent

endowment process. In this model, the agent’s time preference and risk aversion parameters are preselected and the parameters of the endowment process are forced to adapt.

The data are twenty years of monthly observations on consumption growth per capita c_t/c_{t-1} and gross real equity returns r_t^e from January 1959 through December 1978; due to lags $t = 1, \dots, n = 238$. Consumption c_t is nondurables and services per capita and gross equity returns r_t^e are deflated value weighted NYSE returns. These data were provided by the authors Hansen and Singleton (1982), wherein original sources are listed, and are republished in Gallant (1987a), from which the values used here are taken. Some sample characteristics of the data are reported in Table 1.

Table 1 about here

We first calibrate the endowment process by calibrating

$$\log(c_t/c_{t-1}) = a_0 + a_1 \log(c_{t-1}/c_{t-2}) + \sigma_c z_t$$

to current and lagged consumption growth. To relate to previous definitions, $y_t = (c_t/c_{t-1}, c_{t-1}/c_{t-2})$ and $F(y|\rho)$ has parameters $\rho = (a_0, a_1, \sigma_c)$. With no smoothing, the results are $a_0 = 0.0025785$, $a_1 = -0.25929$, $\sigma_c = 0.004160$, and $s_n(\hat{\rho}_n) = 0.050070$. Bootstrapped interquartile ranges for these calibrations are 0.000412, 0.10255, and 0.000277, respectively. With smoothing by the quadratic squasher with $\sigma = 0.010$, results are $a_0 = 0.0026005$, $a_1 = -0.25911$, $\sigma_c = 0.004277$, and $s_n(\hat{\rho}_n) = 0.000010686$. Interquartile ranges for these calibrations are 0.000372, 0.11503, and 0.000268, respectively.

Experience to date with edf-calibration is that for simulation methods that generate a truly continuous distribution, there is little difference between calibrations with and without smoothing as just seen above. But the situation changes when the simulation introduces too much granularity. The smoothing accomplished by the L_2 norm is not enough to overcome it and additional smoothing by the quadratic squasher is essential for stability. We seem to be in that situation with the remaining computations, so we shall only compute calibrations with smoothing hereafter.

Because a first order autoregression fits consumption data reasonably well, it is of interest to compare the direct calibration above to least squares estimates to see if results roughly agree when the model actually fits the data. Unlike edf-calibration, least squares is not invariant to change of scale of measurement. It makes a difference whether one fits directly to $\log(c_t/c_{t-1})$ or to (c_t/c_{t-1}) . Fitting to the former yields $a_0 = 0.002547$, $a_1 = -0.259080$, $\sigma_c = 0.0043549$. Standard errors are 0.00030871, 0.06235477, and 0.000414. The values compare well.

Actually, they must compare well, at least in large samples, because Theorem 1 implies that the calibration is a consistent statistical procedure when $F(y)$ is in the family $F(y|\rho)$, $\rho \in R$. For correctly specified models, edf-calibration must give approximately the same results in large samples as any consistent statistical procedure.

Let p_t be the price of the equity security and presume that the endowment e_t is the dividend paid to the equity asset. The price dividend ratio is then $v_t = p_t/e_t$. Expressed in terms of the price dividend ratio with utility parameterized as

$$\mathcal{E}_0 \left(\sum_{t=0}^{\infty} \beta^t \frac{e_t^{1-\gamma} - 1}{1-\gamma} \right),$$

the asset pricing equation is

$$v_t = \beta \mathcal{E}_t \left[(1 + v_{t+1}) (e_{t+1}/e_t)^{1-\gamma} \right],$$

where \mathcal{E}_t denotes conditional expectation with respect to the endowment process e_t and v_t is viewed as a function of the endowment process. Once v_t is determined, the gross equity return is $r_t^e = [(1 + v_t)/v_{t-1}](e_t/e_{t-1})$ and the gross return on a risk free discount bond is $r_t^b = [\beta \mathcal{E}_t(e^{t+1}/e_t)^{-\gamma}]^{-1}$. Assume that the endowment process $\{e_t\}$ follows the autoregressive process

$$\log(e_t/e_{t-1}) = \alpha_0 + \alpha_1 \log(e_{t-1}/e_{t-2}) + \sigma_e u_t,$$

where $\{u_t\}$ is an uncorrelated Gaussian process.

The price dividend ratio and bond return can be computed using the the method proposed by Tauchen and Hussey (1991). Their method relies on a quadrature based Markov chain approximation to the transition density of the endowment process. For this, we used Tauchen and Hussey's code as distributed by anonymous ftp at ftp.econ.duke.edu in directory pub/get.

The remaining computations are straightforward matrix algebra, which they describe. We chose the number of quadrature points by adding points until the value of $s_n(\hat{\rho}_n)$ stabilized when computed with $N = 50,000$ and $\sigma = 0.01$ for $F_{\sigma,N}(y|\rho)$. Sixteen quadrature points were selected, which allows for $256 = (16)^2$ different equity returns r_t^e . In all computations below, $N = 50,000$ and $\sigma = 0.01$.

We now calibrate Model 1 to gross equity returns for a menu of values for β . In this calibration, we accept the values a_0 , a_1 , and σ_c determined above as the values for α_0 , α_1 , σ_e . To relate to previous definitions, $F(y|\rho)$ for Model 1 has $y_t = r_t^e$ and $\rho = \gamma$. Table 2 displays the results. The corresponding bond and equity returns are shown in Table 3, which may be compared with Table 1. Table 3 presents a rather unattractive menu of parameter choices because the implications for returns are counterfactual

Table 2 about here

Table 3 about here

As seen from Figure 4, the poor match to equity returns is caused more by mismatched scale than by mismatched location. Accordingly, in the next specification we shall calibrate the scale parameter σ_e of the endowment process directly rather than accepting σ_c as its value as above. This can be viewed as a device for adopting the Kydland and Prescott (1996) recommendation that emphasis be placed on matching to location. It also causes endowment to be a partially specified latent process which precludes the use of generalized method of moments.

Figure 4 about here

As before, Model 2 is calibrated to gross equity returns for a menu of values for β ; the values a_0 and a_1 determined from consumption growth are accepted as the values for α_0 and α_1 in the endowment equation whereas σ_e is to be recalibrated. To relate to previous definitions, $F(y|\rho)$ for Model 2 has $y_t = r_t^e$ and $\rho = (\gamma, \sigma_e)$. Table 4 displays the results. The corresponding returns are shown in Table 5. The first line of Table 5 is a plausible specification of preferences with plausible implications for returns. The real return for equities is between the mean and median value shown in Table 1. The ex-ante real bond return is somewhat too high when contrasted with estimates of ex-ante real bond returns reported in the literature, which fluctuate about 2%. Comparing σ_c from Table 2 with σ_e from Table 4 indicates that scale has increased by a factor of 4.5. A visual display of the fit is Figure 3. The contrast with Figure 4 is dramatic.

Table 4 about here

Table 5 about here

The cause of the low values of γ in Table 4 is the high growth rate of the endowment implied by putting $\alpha_0 = a_0 = 0.0026004$. Plausible values for γ over the entire menu of values for β will require lower growth rates. To see this, in Model 3 we shall preselect the agent's utility function parameters and force the endowment process to adapt. Specifically, the model is calibrated to gross equity returns for a menu of values for β with the risk aversion parameter fixed at $\gamma = 1.1$. In this calibration, we accept the value a_1 determined for consumption above as the value for α_1 in the endowment equation, but re-calibrate α_0 and σ_e . To relate to previous definitions, $F(y|\rho)$ for Model 1 has $y_t = r_t^e$ and $\rho = (\alpha_0, \sigma_e)$. Table 6 displays the results. For this calibration, implied returns do not vary across the rows of Table 6. The bond return is 3.102 with a standard deviation of 3.772 and the equity return is 5.015 with a standard deviation of 13.606.

Table 6 about here

To summarize the results, consumption, represented as a first order autoregression in log differences, does a poor job of serving as a single factor for equity returns at reasonable values of the agent's preference parameters for two reasons: the scale parameter is too small and the location parameter (growth rate) is too large. Allowing the scale parameter to adjust freely allows for some plausible values for preference parameters and implied returns. Allowing both to adjust allows preference parameters to be chosen at will and thereby exhibit the characteristics of the single factor implied by that choice.

6 Concluding Remarks

The paper has proposed a calibration strategy that is an improvement on presently available methods in two respects: it has a more compelling rationale and it is safer in that it does not rely on moments that could well not exist. It seems to give reasonable results in applications. Listed below are some loose ends that might be worth attending to in future work.

Matching all possible in-sample probability statements has a stronger rationale than just matching distribution functions. For bivariate data the sample objective function would become

$$s_n^*(\rho) = \frac{1}{n^2} \sum_{s=1}^n \sum_{t=1}^n \left[\hat{P}_n\{(y_s, y_t)\} - P\{(y_s, y_t) \mid \rho\} \right]^2,$$

where

$$P\{(y_s, y_t)\} = F(y_{1t}, y_{2t}) - F(y_{1s}, y_{2t}) - F(y_{1t}, y_{2s}) + F(y_{1s}, y_{2s}).$$

The generalization of this formula to higher dimensions is given by Tucker (1967, p. 24). Among other things, the s_n^* calibration is invariant to a larger class of transformations than the s_n calibration of Section 2. In particular the s_n^* calibration is invariant to a change of sign of one or more variables whereas the s_n calibration is not. The asymptotics of the s_n calibration derived in Section 3 would appear carry over directly to the s_n^* calibration with the obvious change to the population objective function. But computational complexity increases considerably. It is not clear that the benefits are worth the cost.

How large N should be for $F_N(y|\rho)$ to approximate $F(y|\rho)$ adequately has not been discussed. Practically speaking, this is just a computational issue that is easily settled in a given application with a little experimentation. More relevant, as we have seen, is the quality of the simulations, not their abundance. Nonetheless, it should not be too difficult to apply Arcones and Yu's (1994) uniform central limit results and get a probabilistic bound on accuracy that holds uniformly for $\rho \in R$. Much more of a technical challenge would be trying to adapt Yoshihara's (1976) law of the iterated logarithm to get a uniform almost sure bound on accuracy. But while interesting technically, these refinements are not of much practical importance, as just noted.

As noted in Section 1, the proposed calibration is too objective for some tastes. It might be useful to try to derive schemes for blunting its objectivity and sharpening its focus on certain functionals of F . Section 5 contains some experimentation along these lines using unobserved state variables.

A related approach to sharpening the focus of a calibration in a simulation context is the adaptation of notions of encompassing and semiparametric specifications to indirect inference for which see Dridi and Renault (1999) and the references therein. There is some distance between that literature and the considerations here because the encompassing literature is more concerned either with partially specified models or with fully specified models where only part of the specification is asserted to be correct whereas here we are concerned with fully specified models that are not necessarily asserted to be correct in any particular. Nonetheless, there are some points of contact. In the vocabulary of the encompassing literature, what has been proposed here is that the edf proximity function $s_n(\rho)$ replace more customary proximity functions such as Kullback-Liebler discrepancy. Indirect inference requires very simple proximity functions to be computationally feasible. Therefore, while incorporating the edf proximity function into indirect inference would be interesting and feasible theoretically, it is not presently feasible computationally. Indirect inference can be recast into the EMM score based methodology, which permits more elaborate proximity functions, but the edf score would have to be computed by numerical differentiation which would degrade numerical stability to the point of impracticality. Further research might find computational methods that avoid these problems in which case the edf proximity function might prove useful in the

simulation based encompassing literature.

One can calibrate along the same lines as edf-calibration using the empirical characteristic function instead of the empirical distribution function; see Carrasco, Chernov, Florenss, and Ghysels (2001) and the references therein. Matching to the Fourier transform of the distribution function does not seem a natural metric for calibration. It does not satisfy the rationale of fitting to basis functions or probabilities advanced in Section 1. But there is an exception. As shown by Duffie, Pan, and Singleton (2000), the Fourier transform is quite useful for pricing options on affine diffusions. Affine diffusions are rejected as models for both equities and bonds (Chernov, Gallant, Ghysels, Tauchen 2001; Ahn, Dittmar, and Gallant, 2001) so that estimating these models can be regarded as calibration. Matching characteristic functions does seem to be a reasonable metric for calibrating affine diffusions for the purpose of asset pricing.

7 References

- Ahn, D-H., R. F. Dittmar, and A. R. Gallant (2001) “Quadratic Term Structure Models: Theory and Evidence,” *The Review of Financial Studies*, forthcoming.
- Arcones, M. A., and B. Yu (1994), “Central Limit Theorems for Empirical and U -Processes of Stationary Mixing Sequences,” *Journal of Theoretical Probability* 7, 47–71.
- Carrasco, M., M. Chernov, J-P. Florenss, and E. Ghysels (2001), “Estimating Diffusions with a Continuum of Moment Conditions,” Manuscript, Department of Economics, University of North Carolina.
- Caswel, H. (1988) “Theory and Models in Ecology: A Different Perspective,” *Ecological Modelling* 43, 33–44.
- Chernov, M., A. R. Gallant, E. Ghysels and G. Tauchen (2001) “Alternative Models for Stock Price Dynamics,” Manuscript, School of Business, Columbia University.
- Davidson, J. (1994), *Stochastic Limit Theory*, Oxford University Press, Oxford, U.K.

- Doraszelski, U., and Matzkin, R. (2001), “A Semiparametric Estimator for Simultaneous Equations Models,” Manuscript, Department of Economics, Northwestern University.
- Doukhan, P. (1994), *Mixing Properties and Examples*, Springer-Verlag, New York.
- Duffie, D., J. Pan, and K. Singleton (2000), “Transform Analysis and Option Pricing for Affine Jump-Diffusions,” *Econometrica* 68, 929–952.
- Duffie, D. and K. J. Singleton (1993) “Simulated Moments Estimation of Markov Models of Asset Prices,” *Econometrica*, 61, 929–952.
- Durham, G. B., and A. R. Gallant (2001), “Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes,” *Journal of Business and Economic Statistics*“, forthcoming.
- Dridi, R., and E. Renault (1999), “Semiparametric Indirect Inference,” Manuscript, Department of Economics, University of Montreal.
- Eraker, B. (2001), “MCMC Analysis of Diffusion Models with Application to Finance,” *Journal of Business and Economic Statistics* 19, 177–191.
- Huber, P. J. (1981), *Robust Statistics*, Wiley, New York.
- Gallant, A. R. (1987a), *Nonlinear Statistical Models*, Wiley, New York.
- Gallant, A. R. (1987b), “Identification and Consistency in Semiparametric Regression,” in Bewley, T. F., ed. (1987), *Advances in Econometrics Fifth World Congress, Volume 1*, Cambridge University Press, New York, 145–170.
- Gallant, A. R. and G. Tauchen (1996) “Which Moments to Match?” *Econometric Theory* 12, 657–681.
- Gallant, A. R., and G. Tauchen (2001), “Simulated Score Methods and Indirect Inference for Continuous Time Models,” in L. P. Hansen and Y. Aït-Sahalia, eds., *Handbook of Financial Econometrics*, North Holland, Amsterdam.

- Gallant, A. R., and H. L. White, Jr. (1987), *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, Basil Blackwell, Oxford, UK.
- Geweke, J. (1999), "Computational Experiments and Reality," Manuscript, Department of Economics, University of Iowa.
- Gourieroux, C., A. Monfort and E. Renault (1993) "Indirect Inference," *Journal of Applied Econometrics* 8, S85–S118.
- Hall, C. A. S. (1988), "An Assessment of Several of the Historically Most Influential Theoretical Models Used in Ecology and of the Data Provided in Their Support," *Ecological Modelling* 43, 5–31.
- Hansen, L. P., and J. J. Heckman (1996), "The Empirical Foundations of Calibration," *The Journal of Economic Perspectives* 10, 87–104.
- Hansen, L. P., and K. J. Singleton (1982), "Generalized Instrumental Variable Estimation Estimators of Nonlinear Rational Expectations Models," *Econometrica* 50, 1269–1286.
- Hansen, L. P., and K. J. Singleton (1983), "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns," *Journal of Political Economy* 91, 249–265.
- Hansen, L. P., and K. J. Singleton (1984), "Errata," *Econometrica* 52, 267–268.
- Kydland, F. E., and E. C. Prescott (1996), "The Computational Experiment: An Econometric Tool," *The Journal of Economic Perspectives* 10, 69–86.
- Lucas, R. E., Jr. (1978), "Asset Prices in and Exchange Economy," *Econometrica* 46, 1429–1445.
- Manski, C. F. (1983), "Closest Empirical Distribution Function Estimation," *Econometrica* 51, 305–320.
- Matzkin, R. (2001), "Closest Kernel Distribution Estimation," Manuscript, Department of Economics, Northwestern University.

- Mehra, R., and E. Prescott (1985), "The Equity Premium: A Puzzle," *Journal of Monetary Economics* 15, 145–162.
- Müller, U. A., M. M. Dacorogna, and O. V. Pictet (1998), "Heavy Tails in High-Frequency Financial Data," in R. J. Adler, R. E. Feldman, and M. S. Taqqu, eds., *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, Birkhäuser, Boston.
- Nobel, A. B., and R. A. Olshen (1996), "Termination and Continuity of Greedy Growing for Tree-Structured Vector Quantizers," *IEEE Transactions on Information Theory* 42, 191–205.
- Parr, W. C., and T. De Wet (1981), "On Minimum Cramer-Von Mises Norm Parameter Estimation," *Communications in Statistics, Part A, Theory and Methods* 10, 1149–1166.
- Pollard, D. (1984), *Convergence of Stochastic Processes*, Springer-Verlag, New York.
- Politis, D. N., and J. P. Romano (1994), "The Stationary Bootstrap," *Journal of the American Statistical Association* 89, 1303–1313.
- Pötscher, B. M. and Prucha, I.R. (1996), *Dynamic Nonlinear Econometric Models*, Springer-Verlag, New York.
- Sims, C. A., (1996), "Macroeconomics and Methodology," *The Journal of Economic Perspectives* 10, 105–120.
- Tauchen, G., and R. Hussey (1991), "Quadrature-Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models," *Econometrica* 59, 371–396.
- Tucker, H. G. (1967), "A Graduate Course in Probability," Academic Press, New York.
- Withers, C. S. (1981), "Conditions for Linear Processes to be Strong-Mixing," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57, 477–480.
- Yoshihara, K. (1976), "Limiting Behavior of U-statistics for Stationary, Absolute Regular Processes," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 35, 237–252.

Tables and Figures

Table 1. Characteristics of the Data.

Quantiles	c_t/c_{t-1}	r_t^e	$\log(c_t/c_{t-1})$	$\log(r_t^e)$
99%	1.013425	1.121526	0.013336	0.11469
95%	1.009645	1.057568	0.009598	0.055972
90%	1.007642	1.048265	0.007613	0.047136
75% Q3	1.004783	1.030098	0.004772	0.029654
50% Med	1.002235	1.005685	0.002232	0.005669
25% Q1	0.999164	0.978847	-0.00084	-0.02138
10%	0.996466	0.948152	-0.00354	-0.05324
5%	0.994561	0.932595	-0.00545	-0.06978
1%	0.991563	0.890846	-0.00847	-0.11558
IQR Q3-Q1	0.005619	0.051251	0.005608	0.051034
Mean	1.002060	1.002851	0.002048	0.001965
Std. Dev.	0.004573	0.042095	0.004564	0.042143
Annualized				
Med	2.715	7.039		
IQR	1.946	17.754		
Mean	2.500	3.475		
Std. Dev.	1.584	14.582		

Notes: The sampling frequency is monthly. Med is the median and IQR is the inter quartile range. For the normal distribution with mean μ and variance σ^2 , Med = μ and IQR = 1.3490σ .

Table 2. Calibrated Parameters for Model 1

Time Preference		Parameter Calibration Data				
Annual Rate	Monthly β	Equity Returns		Consumption Growth		
		γ	$s_n(\hat{\rho}_n)$	a_0	a_1	σ_c
1%	0.99917	3.16105 (2.7462)	3.385334	0.0026004 (0.000372)	-0.2591061 (0.115025)	0.0042770 (0.000268)
2%	0.99835	2.64983	3.462604	0.0026004	-0.2591061	0.0042770
3%	0.99754	2.12672	3.532321	0.0026004	-0.2591061	0.0042770
4%	0.99674	1.59152	3.593626	0.0026004	-0.2591061	0.0042770
5%	0.99594	1.03804 (3.2180)	3.646292	0.0026004 (0.000372)	-0.2591061 (0.115025)	0.0042770 (0.000268)

Notes: The numbers in parentheses are bootstrapped interquartile ranges based on 1,000 repetitions. For the normal distribution with mean μ and variance σ^2 , Med = μ and IQR = 1.3490σ . The empirical distribution functions of the data and simulation were smoothed with the quadratic squasher with $\sigma = 0.01$ prior to comparison. The simulation size is $N = 50,000$. Values shown for consumption growth are from a separate calibration reported in Section 5.

Table 3. Implied Returns for Model 1

Time Preference		Risk	Annualized Returns			
Annual Rate	Monthly β	Aversion γ	Bond	Std.Dev.	Equity	Std.Dev.
1%	0.99917	3.16105	9.120	1.264	9.232	2.498
2%	0.99835	2.64983	8.846	1.060	8.933	2.261
3%	0.99754	2.12672	8.524	0.850	8.588	2.024
4%	0.99674	1.59152	8.152	0.636	8.196	1.789
5%	0.99594	1.03804	7.727	0.415	7.753	1.558

Notes: The returns were calculated from a simulation with $N = 50,000$ of Model 1 at the parameter settings shown in Table 2.

Table 4. Calibrated Parameters for Model 2

Time Preference		Parameter Calibration Data				
Annual Rate	Monthly β	Equity Returns			Consumption Growth	
		γ	σ_e	$s_n(\hat{\rho}_n)$	a_0	a_1
1%	0.99917	1.24340 (1.8785)	0.0355053 (0.018405)	0.019484	0.0026004 (0.000372)	-0.2591061 (0.115025)
2%	0.99835	0.79303	0.0399196	0.019750	0.0026004	-0.2591061
3%	0.99754	0.46847	0.0435702	0.018653	0.0026004	-0.2591061
4%	0.99674	0.21924	0.0465804	0.017875	0.0026004	-0.2591061
5%	0.99594	0.00625 (0.9175)	0.0492922 (0.014159)	0.017543	0.0026004 (0.000372)	-0.2591061 (0.115025)

Notes: The numbers in parentheses are bootstrapped interquartile ranges based on 1,000 repetitions. For the normal distribution with mean μ and variance σ^2 , Med = μ and IQR = 1.3490σ . The empirical distribution functions of the data and simulation were smoothed with the quadratic squasher with $\sigma = 0.01$ prior to comparison. The simulation size is $N = 50,000$. Values shown for consumption growth are from a separate calibration reported in Section 5.

Table 5. Implied Returns for Model 2

Time Preference		Risk	Annualized Returns			
Annual Rate	Monthly β	Aversion γ	Bond	Std.Dev.	Equity	Std.Dev.
1%	0.99917	1.24340	3.035	4.109	5.102	13.604
2%	0.99835	0.79303	3.445	2.948	4.969	13.610
3%	0.99754	0.46847	3.960	1.901	4.963	13.620
4%	0.99674	0.21924	4.503	0.956	5.018	13.622
5%	0.99594	0.00625	5.019	0.029	5.044	13.623

Notes: The returns were calculated from a simulation with $N = 50,000$ of Model 2 at the parameter settings shown in Table 4.

Table 6. Calibrated Parameters for Model 3

Time Preference		Risk	Parameter Calibration Data			
Annual Rate	Monthly β	Aversion γ	Equity Returns			Consump.
			α_0	σ_e	$s_n(\hat{\rho}_n)$	a_1
1%	0.99917	1.10000	0.0028380 (0.004248)	0.036833 (0.00281)	0.019771	-0.2591061 (0.115025)
2%	0.99835	1.10000	0.0018983	0.036834	0.019771	-0.2591061
3%	0.99754	1.10000	0.0009692	0.036834	0.019771	-0.2591061
4%	0.99674	1.10000	0.0000508	0.036835	0.019771	-0.2591061
5%	0.99594	1.10000	-0.0008682 (0.004248)	0.036835 (0.00281)	0.019771	-0.2591061 (0.115025)

Notes: The numbers in parentheses are bootstrapped interquartile ranges based on 1,000 repetitions. For the normal distribution with mean μ and variance σ^2 , Med = μ and IQR = 1.3490σ . The empirical distribution functions of the data and simulation were smoothed with the quadratic squasher with $\sigma = 0.01$ prior to comparison. The simulation size is $N = 50,000$. Values shown for consumption growth are from a separate calibration reported in Section 5.

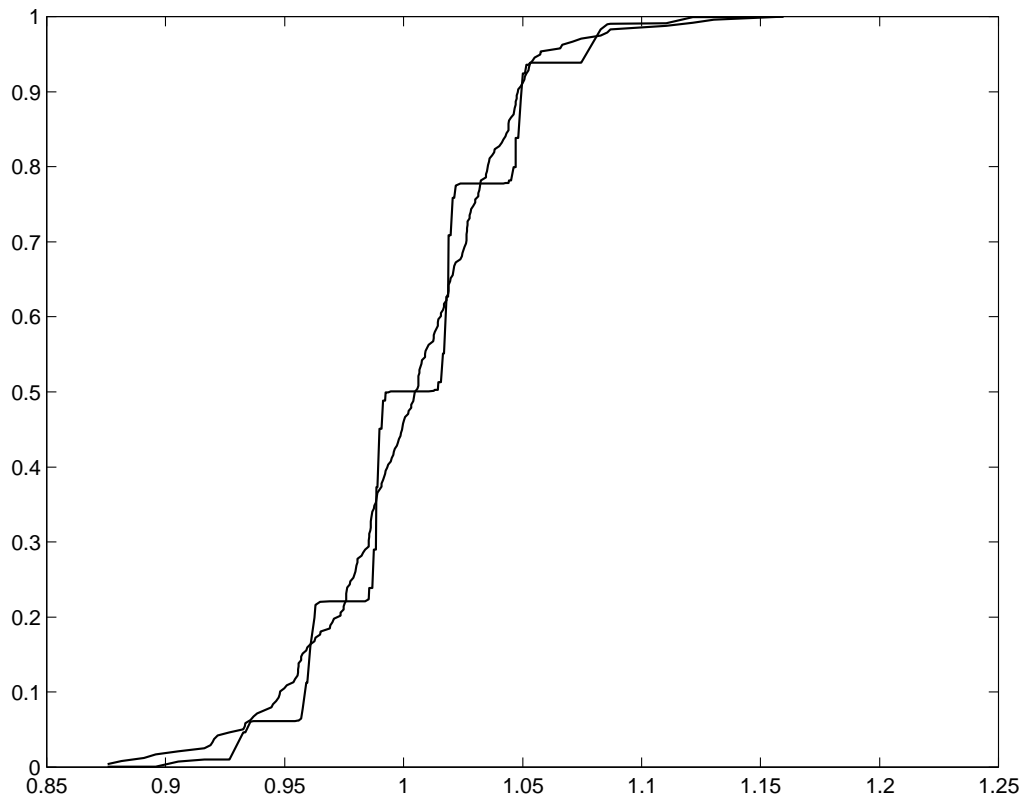


Figure 1. Empirical Distribution Functions of Data and Model. The smooth line is the empirical distribution function of the equity returns data. The stepped line is the empirical distribution function of a simulation from the model at parameter values set to those shown in the first row of Table 4.

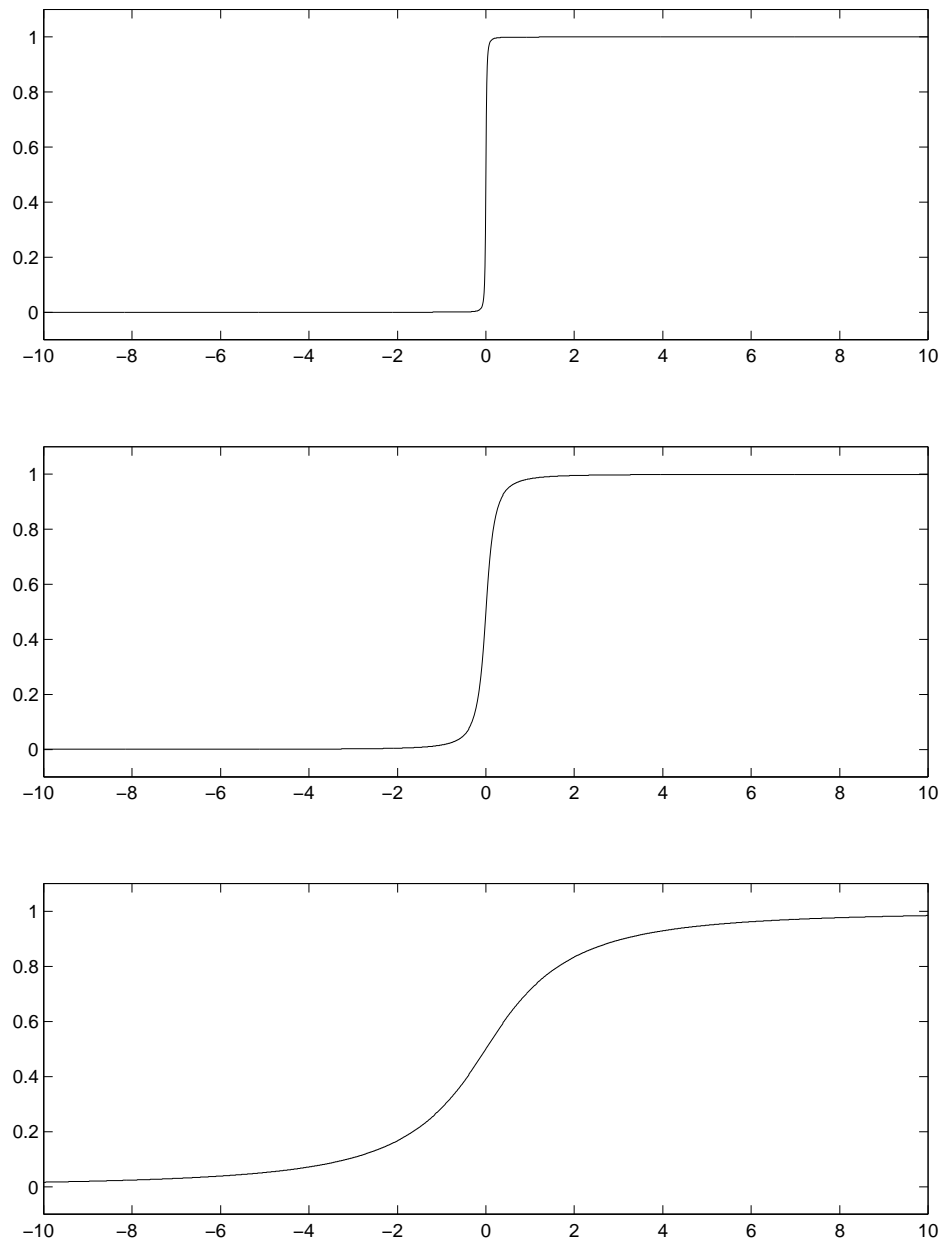


Figure 2. Quadratic Squashers. The top panel plots the quadratic squasher for $\sigma = 0.01$. The middle panel is for $\sigma = 0.1$ and the bottom for $\sigma = 1$.

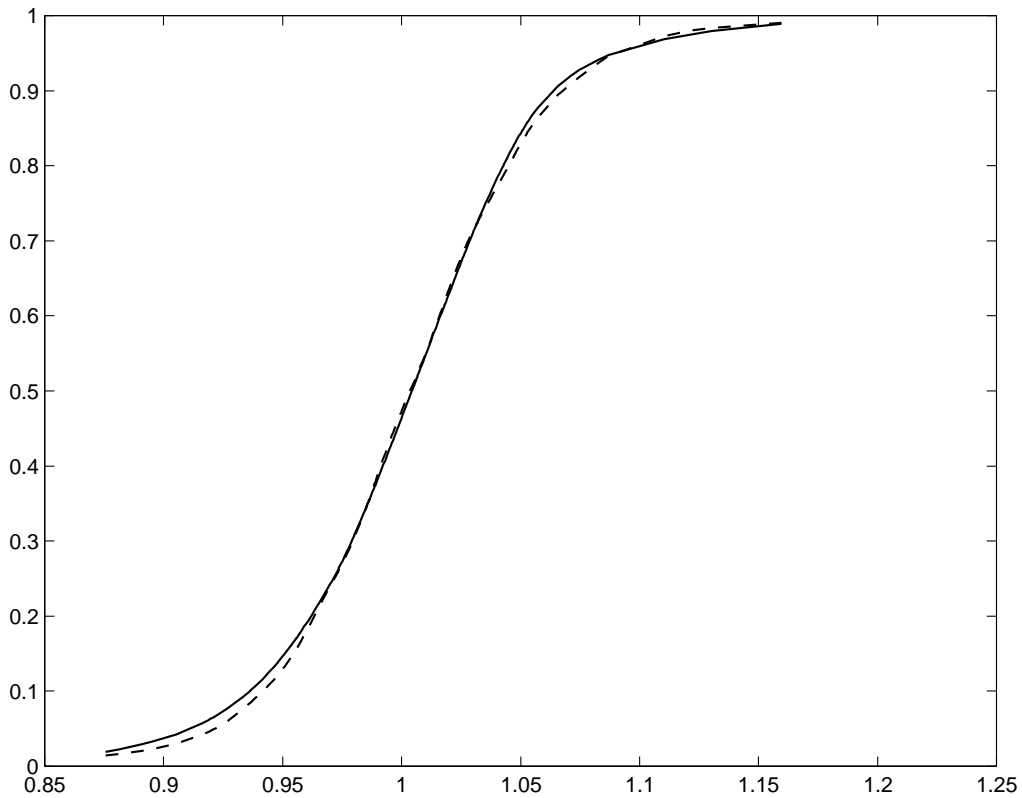


Figure 3. Smoothed Empirical Distribution Functions, Data and Model. The solid line is the empirical distribution function of the equity returns data. The dashed line is the empirical distribution function of a simulation from the model at the parameter values set to those shown in the first row of Table 4. The smoothing shown is accomplished by substituting the quadratic squasher with $\sigma = 0.01$ for the indicator function in the definition of an empirical distribution function.

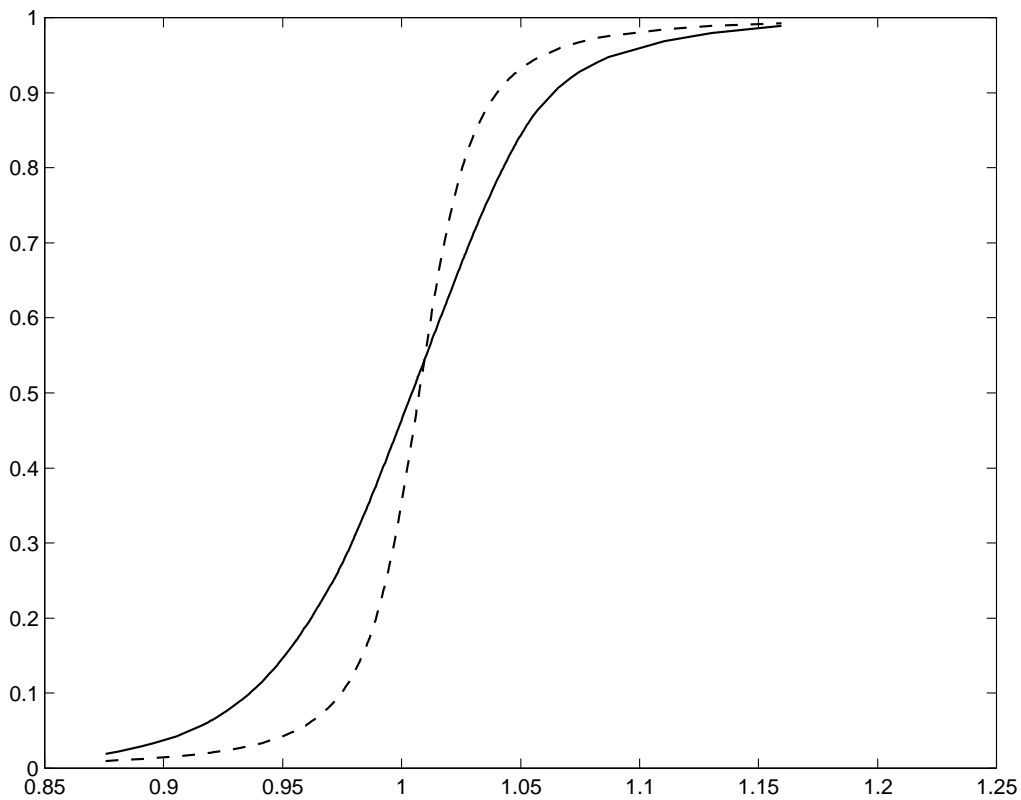


Figure 4. Smoothed Empirical Distribution Functions, Data and Model.

The solid line is the empirical distribution function of the equity returns data. The dashed line is the empirical distribution function of a simulation from the model at the parameter values set to those shown in the first row of Table 2. The smoothing shown is accomplished by substituting the quadratic squasher with $\sigma = 0.01$ for the indicator function in the definition of an empirical distribution function.