

Programm
Surrogator

von

Dr. rer. nat. Hans Uhlig

Copyright Hinweis

Der Text und die Abbildungen dieses Beitrages sowie die zugehörige Software unterliegen dem Urheberrechtsschutz. Wer diese Produkte erwirbt, darf sie für den eigenen Gebrauch nutzen. Kopien oder Wiedergaben in anderer Form ob vollständig oder nur teilweise, bedürfen meiner schriftlichen Zustimmung.

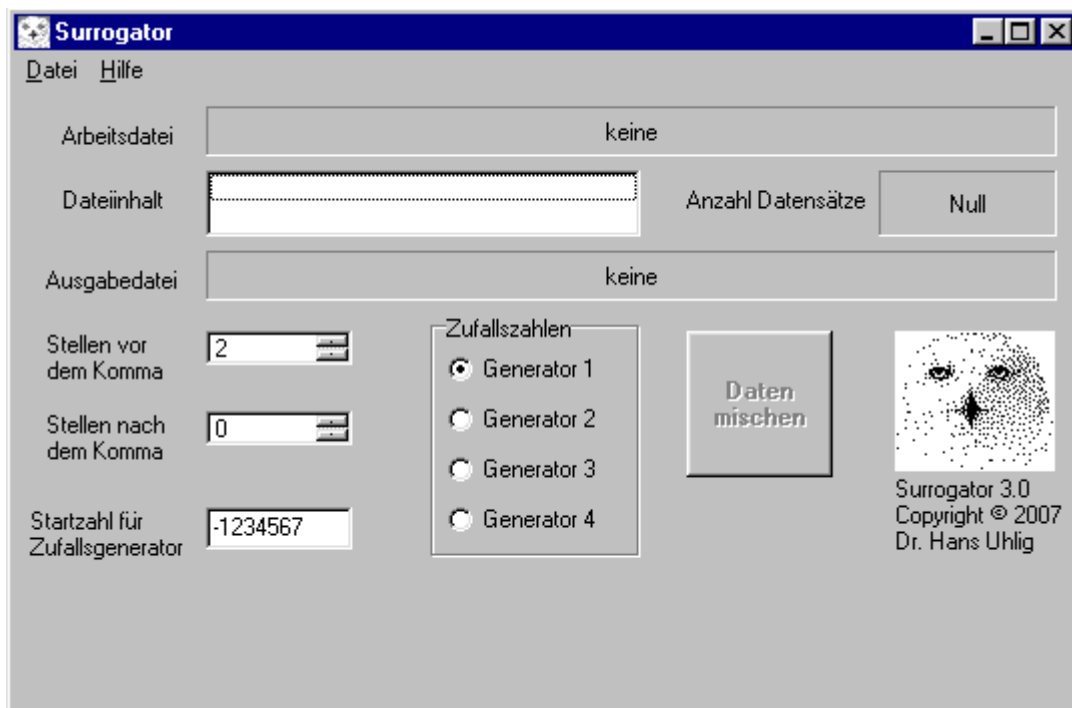
Hans Uhlig, Oktober 2009

Das Programm SURROGATOR

Warum sollte man den SURROGATOR überhaupt verwenden?

Wenn Sie eine Zeitreihe auf Zufälligkeit prüfen, können Sie das mit einem χ^2 -Test auf Unabhängigkeit, durch Testen der Informations-Entropie oder der Transinformation herausfinden. Aber wenn Sie wissen möchten, wie eine zufällig erzeugte Zeitreihe aussehen würde, benötigen Sie eine geeignete Datenserie, die als Null-Hypothese (die Daten sind ungeordnet) dienen könnte. Der SURROGATOR kann diese Daten liefern. Der Vergleich mit den Surrogatdaten, beispielsweise mit dem Programm Entropie-Profil zeigt erst, wie viel Ordnung bzw. Struktur in den Daten enthalten ist. Diese Ordnung kann genutzt werden, weil sie bessere Vorhersagen ermöglicht.

Wenn die wöchentlichen Marktänderungen normalverteilt wären, könnte man eine andere Zeitreihe mit gleicher Verteilung als Nullhypothese heranziehen. Viele (Pseudo)-Zufallszahlengeneratoren können normalverteilte (Pseudo)-Zufallszahlen erzeugen. Dies kann mit dem Kolmogoroff-Smirnoff-Test untersucht werden, der die Methode der Wahl ist, um Verteilungen auf Normalität zu prüfen. Ich habe mit Hilfe dieses Tests viele Märkte untersucht (Uhlig, H., 1999) und in allen Fällen war die Verteilung wöchentlicher Änderungen signifikant, wenn nicht hochsignifikant verschieden von Normalverteilungen. Somit wären Normalverteilungen keine angemessene Nullhypothese. Der SURROGATOR schließt diese Lücke.



Programm Surrogator, Eröffnungsbildschirm. Es wurde noch keine Wertedatei geladen und daher ist auch der Knopf <Daten mischen> nicht aktiv.

Einführung

Dieses Programm läuft auf 32bit Rechnern unter MS-Windows 95,98,2000,XP und folgende. Es erzeugt sogenannte Surrogat-Daten aus Zeitreihen. Der Begriff „Surrogatdaten“ wird für Daten verwendet, die auf unterschiedlichste Weise manipuliert wurden. Darum soll hier klargestellt werden, was damit gemeint ist.

Surrogat-Daten z.B. wöchentlicher Zeitreihen werden durch Zufallstausch der wöchentlichen Änderungen erzeugt. Durch diese Maßnahme bleiben die meisten statistischen Eigenschaften der wöchentlichen Änderungen erhalten: Verteilung, Mittelwert, Median, Modalwert, Varianz, Schiefe und Steilheit sind die gleichen, wie bei der

Originalzeitreihe. Was bei Surrogat-Daten verloren geht, ist die eventuell vorhandene Struktur in der Originalzeitreihe. Dazu gehören lineare und nicht lineare Abhängigkeiten der Daten untereinander.

Programm Surrogator. Es wurde eine Wertedatei geladen, die im Feld ‚Arbeitsdatei‘ angezeigt wird. Der Knopf <Daten mischen> ist aktiv. Das gepunktete Quadrat auf dem Knopf zeigt, dass dieser bereits betätigt wurde. Für die Surrogatdaten wurde ein Dateiname gewählt, der im Feld ‚Ausgabedatei‘ angezeigt wird. Als Datenformat für die Ausgabe wurden zwei Vorkommastellen und drei Nachkommastellen gewählt und die Startzahl für den Zufallszahlengenerator wurde geändert. Für die Herstellung von (Pseudo)-Zufallszahlen wurde der Generator 1 verwendet.

Zufallstausch

Der entscheidende Schritt bei der Erzeugung von Surrogat-Daten ist der zufällige Tausch der Datenreihenfolge. Ein guter (Pseudo)-Zufallszahlengenerator ist dafür zwingend erforderlich. Wir stellen hier gleich mehrere solcher Generatoren zur Verfügung. Es sind Generatoren, die nach verschiedenen Prinzipien arbeiten. Sie verwenden Algorithmen (ran0 bis ran3 – für englisch: ‚random numbers‘ = Zufallszahlen), die in der NUMERICAL RECIPES ^(TM)-Programm-bibliothek verfügbar sind, siehe Hinweis am Ende dieser kurzen Beschreibung.

Der Generator 1 entspricht dem Programm ran0 aus den NUMERICAL RECIPES. Im Gegensatz zu den anderen Generatoren ist er nicht portabel. Das heißt, er verwendet die interne Pseudozufallsfunktion des jeweiligen Rechners, auf dem das Programm läuft. Daher wird das Programm auf verschiedenen Rechnern nicht die gleiche Datenreihenfolge liefern. Da die eingebauten Zufallsfunktionen nicht besonders gute ‚Zufallszahlen‘ liefern, werden diese von ran0 noch mit einem weiteren Mischalgorithmus nachbehandelt mit dem serielle Korrelation aufgehoben wird.

Die Generatoren 2,3 und 4 sind portabel. Es sind die Funktionen ran1, ran2 und ran3 aus den NUMERICAL RECIPES: Sie verwenden eine eigene Funktion zur Erzeugung von Zufallszahlen und können daher auf jedem Rechner die gleiche Datenfolge liefern. Die Generatoren 2 und 3 sind besonders ausgesuchte Generatoren herkömmlichen Typs, während der Generator 4 (ran3) nach einem anderen Prinzip arbeitet. Der Algorithmus wurde ursprünglich von Donald Knuth, dem bekannten Autor der Buchreihe „The art of computing“ (Die Kunst des Programmierens) vorgeschlagen. Anders als die meisten (Pseudo)-Zufallszahlen-Generatoren, die in Programme, wie Compiler oder Kalkulationsprogramme, eingebaut sind, ist er kein sogenannter „linear kongruenter“ Generator. RAN3 verwendet die sogenannte Subtraktionsmethode, um (Pseudo)-Zufallszahlen zu erzeugen. Dieser Generator hat alle Tests mit Bravour bestanden, die vorgeschlagen wurden, um damit erzeugte Zahlenreihen auf „Zufälligkeit“ zu testen. Er hat auch den „Serientest“ absolviert, bei dem etliche linear kongruente Generatoren versagen.

So erzeugt man Surrogat-Daten mit dem SURROGATOR

Daten laden

Das Programm benötigt Daten, die auf einer Datei bereit gestellt werden müssen. Die zu bearbeitende Datei darf, aber muss sich nicht im gleichen Verzeichnis befinden.

Datenanforderung

Das Programm verarbeitet Textdateien, die einen Datenwert je Zeile enthalten. Es darf keine Leerzeile in der Datei enthalten sein, weder am Anfang, noch in der Mitte oder am Ende. Wenn diese Daten Nachkommastellen enthalten, müssen sie, wie im Englischen üblich, mit einem Dezimalpunkt abgetrennt werden. Falls man, wie im deutschen Sprachraum gewohnt, ein Komma zum Abtrennen verwendet, gibt es eine Fehlermeldung.

Denn der Computer erkennt dann die Zeichenfolge nicht als Dezimalzahl, sondern als Text. Wenn dieses Problem auftritt, kann man es leicht beheben, indem man mit einem Texteditor, der reine ASCII-Dateien erzeugen kann (z.B. WordPad), diese Dateien bearbeitet. Man muss nur die Funktion ‚suchen und ersetzen‘ aufrufen und das Komma gegen einen Punkt austauschen. Wenn man im Menüpunkt ‚alles ersetzen‘ wählt, ist die Sache damit erledigt.

Vor dem Datenwert darf keine weitere Zahl, beispielsweise eine Datumsangabe stehen, andernfalls wird das Programm mit einer Fehlermeldung abgebrochen. Solange keine Daten geladen sind, kann man auch keine Surrogatdaten erhalten.

Die Datei darf nicht beliebig lang sein. Das Maximum sind 5000 Daten, was beinahe 100 Jahren an Wochendaten entspricht, also eine Obergrenze festlegt, die praktisch nie erreicht wird.

Zufallsgenerator auswählen

Der Generator 1 ist voreingestellt, aber man kann einen beliebigen Generator auswählen. Dazu muss man den entsprechenden Radiobutton anklicken.

Startwert

Die Zufallsgeneratoren benötigen einen Startwert. Mit dem gleichen Startwert erhält man bei wiederholten Versuchen immer die gleiche Datenfolge. Der Startwert muss eine ganze Zahl sein, negativ oder positiv und von beinahe beliebiger Größe. Voreingestellt ist der Startwert –1234567.

Prinzipiell sind alle Werte (ganze Zahlen) gleich gut. Aber was für eine lange Datenreihe gut ist, muss nicht auch für eine kurze Reihe gut sein. Darum sind in der Praxis nicht alle Werte auch wirklich gleich gut. Da hilft nur ausprobieren. Aber je länger die gewünschte Datenreihe, um so geringer werden die Qualitätsunterschiede.

Kommastellen für Datenausgabe wählen

Die Surrogatdaten sollen in eine Datei geschrieben werden und dafür ist es nötig, anzugeben, welches Format die Daten haben sollen. Viele Aktienmärkte verwenden zwei Nachkommastellen, andere keine und bei Währungen sind vier Nachkommastellen üblich. Die Stellen vor dem Komma sind ebenfalls unterschiedlich: für viele Märkte reichen vier Stellen aus, aber für einige Märkte sind fünf nötig und für den BOVESPA in Brasilien braucht man sogar sechs Vorkommastellen.

Voreingestellt sind 2 Stellen vor dem Komma (bzw. Dezimalpunkt) und 0 Stellen danach. Durch Drücken der Pfeiltasten kann man die Anzeige der Kommastellen ändern. Bitte beachten Sie: Durch die Anzeige allein wählt man die Angaben noch nicht aus. Dies geschieht erst durch das Anklicken der angezeigten Zahlen. Erst nach dem Anklicken wird der Hintergrund blau und die Zahlen sind gewählt.

Daten mischen

Sobald die Datei mit den zu bearbeitenden Daten geladen ist, kann man diese mischen, indem man den entsprechenden Knopf anklickt. Beachten Sie, dass dann die voreingestellten Angaben verwendet werden.

Diese Surrogatdaten, die man nach dem Mischen erhält, besitzen die gleichen Anfangs- und Endwerte, wie die Originaldaten. Nur im langen Mittelteil unterscheiden sie sich. Das Programm geht von der eingelesenen Datei aus. Von aufeinander folgenden Daten werden

Quotienten gebildet, und in eine interne Datenbank eingelesen. Die Reihenfolge der Quotienten wird einem Zufallstausch unterworfen und dann werden die Quotienten in der neuen Zufallsreihenfolge wieder mit dem Anfangswert multipliziert. So erhält man eine neue Datenfolge, die den gleichen Anfangswert und den gleichen Endwert besitzt, wie die Originaldatei. Das ist natürlich zu erwarten, denn es spielt keine Rolle, in welcher Reihenfolge man gegebene Faktoren miteinander multipliziert. Das Endergebnis wird von der Reihenfolge nicht berührt, nur die Zwischenwerte können verschieden sein. In der Regel wird man zum Vergleich mit einer Originalzeitreihe nicht nur eine, sondern mehrere Surrogatzeilen erzeugen, z.B. 5, 10, 100 oder mehr.

Um verschiedene Surrogatdatenreihen zu erzeugen, kann man jeweils den gleichen Generator aber mit verschiedenen Startwerten verwenden, oder man kann den gleichen Startwert mit verschiedenen Generatoren verwenden. Im zweiten Fall sind damit freilich nur vier verschiedene Datenreihen möglich.

Anmerkung

Das eigentliche Mischen der Daten geschieht erst beim Abspeichern in eine Datei, jedoch nur dann, wenn der Knopf <Daten mischen> vorher gedrückt wurde.

Daten abspeichern

Um die Daten abzuspeichern, muss man das Dateimenü öffnen und den entsprechenden Menüpunkt auswählen. Man hat bei der Wahl des Dateinamens, des Namenszusatzes und des Verzeichnisses freie Hand. Falls man vergessen hat, die Daten zu mischen, werden die Daten in der ursprünglichen Reihenfolge abgespeichert, aber mit einem fest vorgegebenen Format von 6 Vorkommastellen und 4 Nachkommastellen. Außerdem erhält man beim Abspeichern die Meldung, dass die Ausgabedatei gleich der Einlesedatei ist.